AFRL-IF-RS-TR-2003-124
Final Technical Report
May 2003

SiGe HBT BiCMOS FOR 2-160 Gb/s NEXT GENERATION INTERNET (NGI)

Rensselaer Polytechnic Institute

Sponsored by
Defense Advanced Research Projects Agency
DARPA Order No. K212

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2003-124 has been reviewed and is approved for publication.

APPROVED:

ROBERT L. KAMINSKI
Project Engineer

FOR THE DIRECTOR:

WARREN H. DEBANY, Jr., Technical Advisor
Information Grid Division
Information Directorate

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 074-0188*

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
May 2003

3. REPORT TYPE AND DATES COVERED
Final Jun 01 – Dec 02

4. TITLE AND SUBTITLE

SiGe HBT BiCMOS for 2-160 Gb/s NEXT GENERATION INTERNET (NGI)

5. FUNDING NUMBERS
C - F30602-01-1-0533
PE - 62302E
PR - K212
TA - NG
WU - I2

6. AUTHOR(S)

John F. McDonald

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Rensselaer Polytechnic Institute
Center for Integrated Electronics
Troy, NY 12181

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Advanced Research Projects Agency
3701 N. Fairfax Drive
Arlington, VA 22203-1714

AFRL/IFG
525 Brooks Road
Rome, NY 13441-4505

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

AFRL-IF-RS-TR-2003-124

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: Robert L. Kaminski, IFG, 315-330-1865, kaminskir@rl.af.mil

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***
This Final Report describes the research related to high speed serial communication circuits implemented in SiGe HBT technology at Rensselaer Polytechnic Institute. Serializer/Deserializer (SERDES) circuits are crucial in keeping pace with the rapidly advancing needs for high-speed data transmission in both short distance and long distance scenarios. Research was undertaken to make better use of existing long-haul infrastructure such as fiber optic networks, as well as for improving communication over much smaller scales such as those distances found on a typical PCB. Designs pushing the fundamental limits of the available manufacturing processes provide a fertile ground for developing innovative circuits in both the analog and digital realms. Work rapidly changes focus to the circuitry responsible for the most recently encountered bottleneck, be it in amplification, digital sampling, oscillators, or elsewhere. To date we have created two complete prototype designs, which were fabricated. A third design would capture 80% of the device fT as a bit rate. For IBM 5HP with an fT of 50 GHz this would be 40 Gb/s, but for a 210 GHz ft in IBM 8HP this would become 160 Gb/s. The challenge is to achieve unprecedented symmetry in the SERDES circuit and layout, as well as development of extremely low jitter VCOf's. Figure 7.15 at the end of this report shows just how good the result may be after all the circuit work is completed.

14. SUBJECT TERMS
Communication circuits, serializer, deserializer, data transmission , SERDES

15. NUMBER OF PAGES
92

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Table of Contents

List of Figures and Tables

Chapter 1 Project Overview

History

This project was initially started in 1998 by Thomas Krawczyk with the goal of developing a monolithic 20Gb/s SiGe serializer/deserializer, (or SERDES), circuit to be used in short haul operations. This project was funded by the Naval Research Labs, (NRL). One of the project goals was to try to retain flexibility in data rates for interoperability, while simultaneously being capable of achieving high maximum data rates. Dr. Jack McDonald was responsible for obtaining the funding for this project and overseeing the work. Peter Curran joined the project full time in late 1999. The major focus of the past work has been participation in the creation of two SERDES designs related to this project, both of which utilized 50GHz SiGe HBTs. These designs were 20 Gb/s parallel to serial (Tx) and serial to parallel (Rx) circuits using almost exclusively fully differential CML. The designs incorporated 8-phase VCOs as elements in Phase 2 Locked Loops (PLLs). Thomas Krawczyk departed early in 2001 after receiving his degree, and Peter Curran continues the project, attempting to increase operating rates towards 40Gb/s in the same technology.

Project Timeline/Description

The SERDES I chips were delivered unpackaged, in wafers, necessitating the use of the probe-station testing lab. These circuits significantly under-performed, primarily due to the inexperience in using parasitic extraction tools. Several design errors were identified, and in Nov 1999 work began on the SERDES II chip. For this design cycle, there was a need for more flexibility and testability. This chip contained significant advances in the area of testability, including the incorporation of an on-chip BER test system. The SERDES II chip design was sent for fabrication in Apr 2000, sponsored by Sierra Monolithics, Inc.

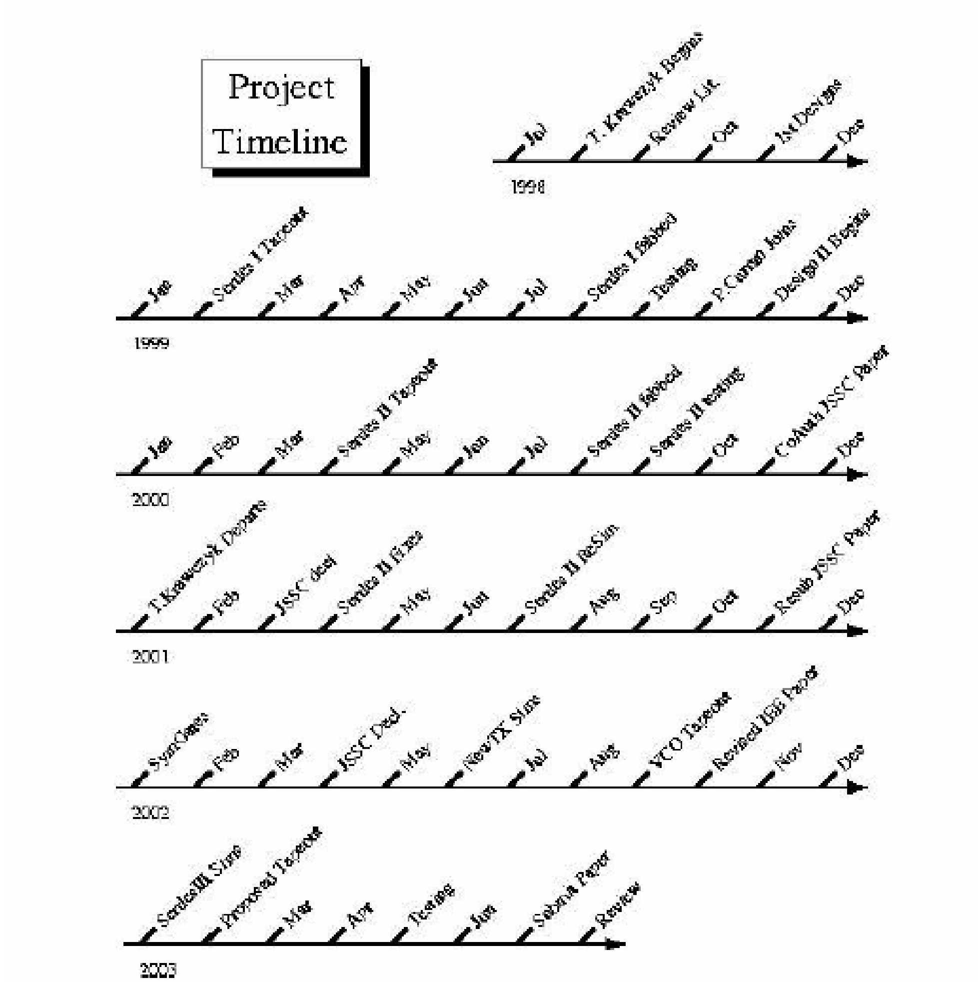

Figure 1.1: Research Timeline

The SERDES II chip was delivered in August 2000. Performance was again tested and deviations from expected behavior analyzed. To facilitate testing, LabVIEW software was written to control the test equipment, including a rented spectral analyzer. A paper describing results was submitted to the JSSC in Nov 2000. Due to the tight fabrication schedule, several errors were incorporated into the SERDES II design. The paper was declined for publication in March 2001, most likely due to these flaws. These were identified and corrected in layouts, and then re-simulated.

Although newer 120GHz f_T SiGe processes are available they are quite expensive. Most if not all of the circuits can be implemented in the newer technologies with a corresponding increase in operating rate.

Chapter 2 SERDES Overview

Serial Communication Systems

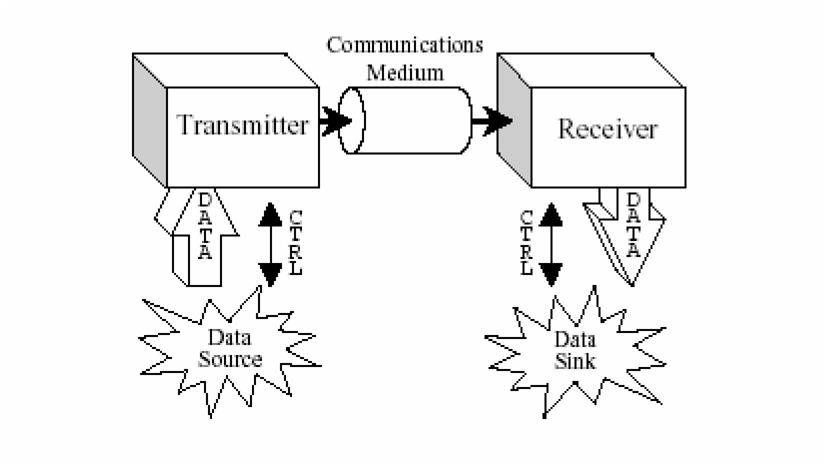

Figure 2.1: Serial Comm. System

Most high-speed serial communications systems in their simplest form can be broken down into 3 major components, (see Figure 2.1). These components are the Transmitter, the Communications Medium, and the Receiver. The Transmitter typically takes in parallel data from a data source and then outputs it serially via multiplexing or via a shift register. The transmitter usually requires some form of additional communication with the data source in order to ensure that correct data is read. The serial data, typically with no additional clocking information, is sent out along the communications medium. Because no separate clock signal is sent, the data itself must be used to recover a clock.

The communication channel consists of some material medium, as well as the electronics necessary to convert incoming/outgoing electrical signals. The actual signals transmitted over the medium may use an entirely different data encoding than the incoming or outgoing electrical signals. PSK, (Phase Shift Keying), is a common encoding. Some channel interfaces might require single-ended signals as opposed to differential. Certain types of channels may put constraints on the outgoing data such as a requirement of having a zero average DC component, necessitating the transmission of as many zeros as ones. Channels may require drivers, receivers, and often repeaters, which influence the final signal that arrives at the Receiver(Rx). Losses along the channel and SNR ultimately limit performance. A single communication channel might be used by different independent serial streams simultaneously, as in the case of WDM(wave division multiplexing.) In the context of this document, the communications medium represents the medium, the repeaters, and drivers necessary to communicate with the transmitter and receiver electrically, and will not be further addressed.

At the Receiver the incoming signal is examined, bit values extracted, and the data is output again in a parallel form. As with the transmitter, the transfer of the parallel data usually requires some form of additional communication with a data sink to transfer the information reliably. The Receiver must accurately recover the data after it has passed through the channel., and to do this it usually has to somehow extract clocking information from the serial data stream.

Transmitter

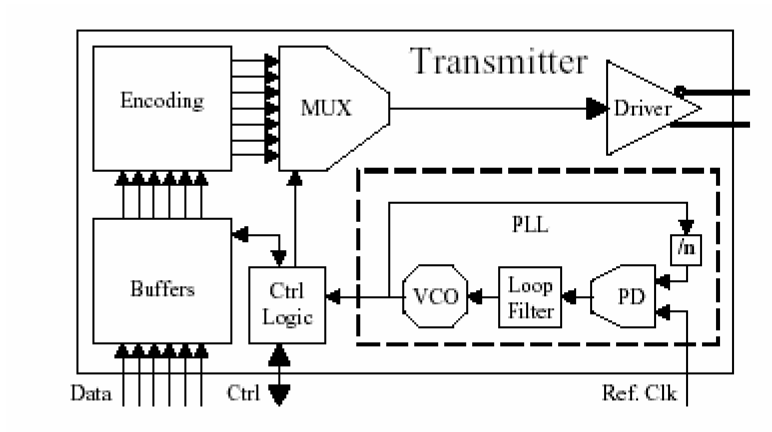

Figure 2.2: Transmitter Block Diagram

Figure 2.2 shows a simplified representation of a Transmitter. There must be some form of communication, (in this case labeled “Ctrl”), between the Transmitter and the source of the parallel data to ensure the data at the inputs is valid when required by the Transmitter. Depending on the application, asynchronous parallel input to the transmitter might be desired and implemented using buffering. Control lines indicating FIFO status are common. The parallel data bits can be from a single source or from multiple sources such as in TDM(time division multiplexing). Encoding before transmission is done on chip when necessary. In certain cases, the communication at each of the parallel lines is itself implemented as a complete SERDES system at a lower data rate. The transmitter requires drivers to send the signals off chip, meeting signal levels and matching impedance characteristics. The data at the transmitter output may be transmitted continuously over the communication medium or it may be broken up into packets. Packets are often used in a shared channel, along with some mechanism for detecting resource-sharing conflicts.

The transmitter is responsible for the accurate timing of it's output, without which the receiver on the far side of the communications medium cannot operate. The transmitter thus requires an accurate reference clock, often generated off chip or using off chip components such as crystals. When the reference clock is generated off chip, the transmitter may also generate a clock signal internally which is synchronized with the external clock using a PLL, (phase locked loop). This internal clock may run at many multiples of the external reference and may also be multi-phase.

Receiver

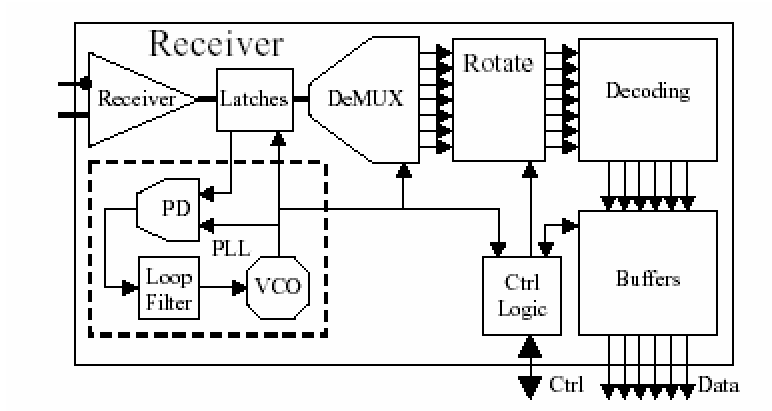

Figure 2.3: Receiver Block Diagram

The Receiver, shown in Figure 2.3, must be able to extract a clock signal from the incoming signal in order to accurately sample the signal's data. This is usually accomplished using a PLL that adjusts the frequency and phase of a local oscillator to match that of the incoming data stream. When the transmitter is not sending data, the receiver cannot synchronize itself with the non-existent incoming signal, and this condition is called "out of lock". After the transmitter begins sending data, a certain amount of time is required for the receiver to regain synchronization, which is termed the lock acquisition time. In some cases trade-offs must be made between acquisition time and locking range. Some packetized systems might make use of multiple transmitters that have completely dis-synchronous phases. Once in lock, rotation and decoding circuitry are often used to prepare the data for output.. The received data must be presented in a parallel form at the outputs along with a reference signal indicating when the data is valid. Buffering may be used to asynchronously pass data between the receiver and the data sink(s). Various indicators of receiver status are also commonly implemented such as loss of lock(LOL), buffer overruns, etc. When the parallel data is intended for different destinations, it is sometimes advantageous to use another lower data rate SERDES system for each of the parallel output lines.

Formatting/Encoding

Typically the transmitter and receiver will use binary voltage levels at their respective inputs and outputs, often sending the signals differentially in the case of high-speed systems. Even in the limited arena of binary voltage signaling, there are multiple ways of representing data bits. As an example, two such methods are RZ and NRZ encoding. RZ (return to zero) encoding uses data bits with a zero bit interspersed between each of them. NRZ (non-return to zero) encoding omits the interspersed zero. The RZ encoding has edges which are closer together for a given bit rate when compared to NRZ. This has obvious negative implications due to the increase in necessary bandwidth, or corresponding reduction in bit rate. The reason RZ encoding is sometimes used is because it allows easier reconstruction of the clock from the transmitted signal as it's spectrum has a peak at the clock frequency. The NRZ encoding has no such peak, and clock recovery can be more complicated.

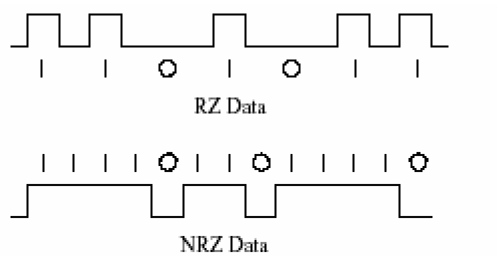

Figure 2.4: Data Formats

Data can also be encoded in ways to ensure data integrity, or to alleviate some problematic property of the data channel. Parity bits and CRCs can be generated to reduce the risk of data corruption. As an example of encoding for a specific property, many communication systems require a DC average value of 0V, necessitating an equal number of ones and zeros. Another common requirement is the minimum density of edges in the data stream. This is analogous to requiring a maximum length sequence of ones or zeros. In addition to being necessary to maintain receiver synchronization, frequent edges are necessary in capacitively coupled systems to prevent droop due to RC decay. Methods for ensuring these properties include Huffmann and 8B/10B encoding.

Framing

The individual bits in a serial stream do not have an inherent property that allows them to be directed to a specific parallel output line. If the receiver can acquire lock at any point in the serial stream, then the parallel output lines can end up "rotated" with respect to the data presented to the transmitter's parallel input lines. Without placing constraints on the data in the serial stream, detection of this rotation is impossible. Depending on the application, there are several ways to correct for this. In a packetized system, a synchronization field can be used with a specific bit pattern that allows the receiver to detect the rotation. A barrel rotation circuit could then automatically correct for this condition before the real data is received. This would require that some framing related encoding scheme be pre-selected and would reduce the generality of the circuit's application. Additionally, a synchronization field would also allow the receiver time to acquire lock, before the actual data payload is received.

In a non-packetized system, an encoding scheme for the data can be chosen which would indicate rotation, perhaps by having one of the bits in the "word" dedicated for such a purpose, or encoded in a detectable manner which uses less of the bandwidth. Encoding the data at the transmitter would involve adding auxiliary data and would reduce the available bandwidth.

Different methods might be more suitable for single source parallel data as opposed to multiple. For that reason, the encoding is often left to a higher level circuit in order to preserve generality. A general-purpose receiver may then incorporate a barrel rotation circuit controlled by an external higher level circuit that detects the rotation and can correct for it.

Serial communications research is leading to rapid advances in data transmission rates. The absolute data rates achieved are not necessarily the primary consideration as cost is a major concern for the implementation of these technologies. Circuits using inexpensive high yield manufacturing processes such as SiGe[1] can be much more desirable than those using slightly faster more exotic III-V semiconductors. For valid comparison, circuit performance should always be qualified by the performance of the devices used in the implementation. SERDES-related research encompasses many different circuit design areas including voltage controlled oscillators(VCOs), phase detectors(PDs), amplifiers, filters, latches, multiplexers, and demultiplexers. They form subsystems of SERDES circuits such as phase locked loops(PLLs), clock multipliers, data retimers, and clock and data recovery units(CDRs). Only when all the individual subsystems have been developed can an integrated monolithic implementation be achieved.

Chapter 3 State of the Art

There are many different categories of research papers that can be relevant to SERDES design. Among these are papers describing multiplexers, demultiplexers, AGC amplifiers, phase locked loops, VCOs, CDRs, as well as complete SERDES implementations. The ratio of data transmission rate to maximum device operating frequency is our primary metric. The research was focused within the regime of monolithic integrated designs. This is dictated primarily by the constraints of the probing station and testing equipment. External loop filter components or crystals cannot be easily connected to the unpackaged chips.

Although SiGe is not the fastest technology, (InP is faster), it is very competitive from a cost point of view, as well as having benefits due to the ability to integrate several subsystems on one die when using a bicmos process. Current high-speed serial research in SiGe is focused on 40Gb/s with 120GHz f_T devices, while this design remains at the 50GHz f_T technology node. Several of the circuits necessary to realize complete SERDES systems at these data rates with 50GHz f_T devices have been fabricated and tested, but many components have not.

The first applicable paper seriously examined when this project began described a complete monolithic10Gb/s TX/RX chipset[2] in a 25GHz f_T Si-Bipolar process, published in 1998, which was used as a launching point for the work. This paper described a method of using multiphase clocks to extract timing information for a high speed receiver. A "leap frog" ring VCO architecture was described, as well as an "f_T doubler" circuit. The paper dealt with some of the issues regarding the generation of the transmitter output data from multiphase sources. One goal of the research was to investigate flexible data rate mechanisms. A multiple data rate CDR circuit is presented in [3], which detects incoming data rate and adjusts a PLL divider automatically. In [4], a 12.5Gb/s Tx/Rx set of PLLs are described using a 45GHz f_T SiGe process. CDRs capable of 19GHz[5] using HEMTs with 50GHz f_T have been reported as early as 1994.

Data rates of 50Gb/s for a simple MUX and 46Gb/s for a DEMUX were reported[6] in a Silicon Bipolar technology with an f_T of 36GHz, along with a 30GHz static frequency divider, in 1996. These circuits were simple E_2CL elements driven by external clocking. Dynamic frequency dividers with rates as high as 79GHz in an 80GHz f_T technology have been reported[7]. The authors report 53GHz static frequency dividers in the same technology.

In [8], a 40Gb/s CDR is described in 50GHz f_T SiGe technology in a non-monolithic approach. A 40Gb/s bit stream is split into two 20Gb/s streams using an external VCO and PLL. Only the phase detector is on-chip. The total on-chip component list consists of 3 flip-flops and an XOR. Amplifiers with AGC for 40Gb/s receivers[9] have been available since 1998 using 92GHz f_T devices.

Technology is rapidly advancing, and 40Gb/s SERDES implementations now exist in SiGe using faster transistors such as 120GHz f_T[10], and a CDR using InP 160GHz f_T devices is described in [11]. The SiGe design uses a CMOS LC quadrature oscillator, while the InP uses a bipolar oscillator using delay lines as the tuning element. A 40Gb/s CDR has also been realized in a 72GHz f_T SiGe HBT technology[12]. This uses a single phase 40GHz VCO implemented with delay lines which is then statically divided to obtain quadrature phases.

Chapter 4 Technology

Initial Choices

Decisions in choosing the process and logic for the project were influenced by both ambition and practicality. Having been given access to IBM's advanced SiGe HBT process made the process choice clear. The only other options would involve expensive low yield compound semiconductor processes. Industry seems to be embracing SiGe because of its obvious benefits. Work by other members of the research group in the FRISC project influenced the decision to go with fully differential CML logic with optional emitter followers. Basic principles surrounding the use of SiGe in HBTs was outlined in [13] in 1982.

IBM SiGe5HP BiCMOS Process

IBM's 5HP BiCMOS process allows the use of 0.5µm x 1.0µm emitter HBTs and 0.35µm L_{eff} CMOS on the same die[1]. It is a fully featured process with numerous types of resistors, and includes components like varactors, capacitors, inductors, and diodes. The logic designs primarily use the npn HBT, and the polysilicon resistor.

The SiGe HBT

IBM's SiGe HBT differs from a normal BJT in that the base contains germanium. The base is epitaxially grown with a non-uniform doping profile. The introduction of germanium into the transistor base results in a reduction of the base bandgap increasing emitter injection efficiency(ϒ) and reducing base transit time. The improvement in injection efficiency would normally result in improved beta, but this is traded off for reduced base resistance by increasing the base doping. With this change, greater currents can be used.

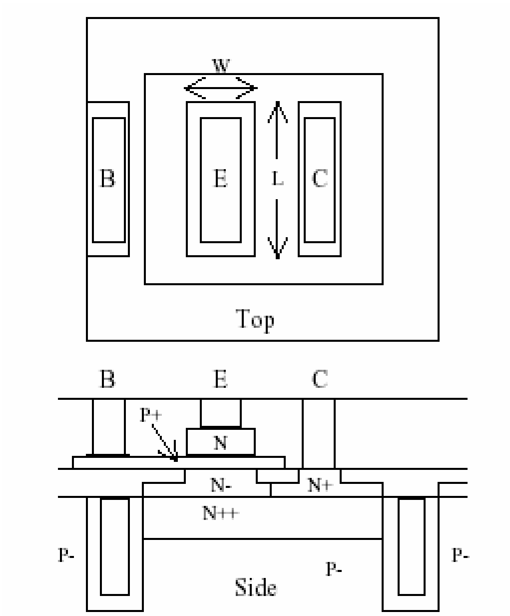

Figure 4.1: Layout of the 5HP SiGe HBT showing top and side views.

In Figure 4.1 above, the general layout of the npn HBT is shown. The layout cell in the design kit is parameterized by the emitter length(L), while the width(W) is fixed at 0.5μm, as indicated in the top view. The emitter and collector areas, (as visible from the top), are surrounded by a deep trench isolation moat that serves to define the N++ subcollector, the collector itself, and the subcollector connection. The P+ SiGe base is epitaxially grown. There are two types of HBTs available, npn, and npnhb. The npnhp is a high breakdown version that has a different collector doping in order to allow higher collector-emitter voltages, up to 4.5V.

The IBM design kits contain subcircuit models for the HBT which contain extrinsic resistances, a VBIC(vertical bipolar intercompany model)[14], and an optional current source to degrade performance appropriately when a high breakdown HBT is instantiated. The design kit is optimized for Cadence's Spectre circuit simulator that includes a VBIC model with parameters for self heating and impact ionization.

This project had access to the design kit through several iterations of model refinement, so over time the device model parameters have varied by several percent. Presented below are some device characteristic curves produced by simulation using the design kit.

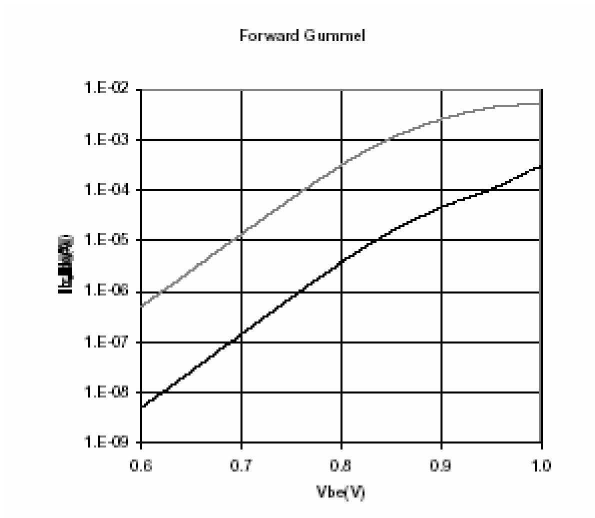

Figure 4.2: Gummel plot of 2.5μm npn HBT at 70C.

Figure 4.2 is a forward active Gummel plot of a 2.5μm npn showing β to be approximately 100 at 70C with V_{CE}=0V. The β varies only slightly with transistor size, and the high breakdown npnhb shows a β of approximately 90. As temperatures go up, so do the absolute currents, but the ratio stays approximately the same as expected.

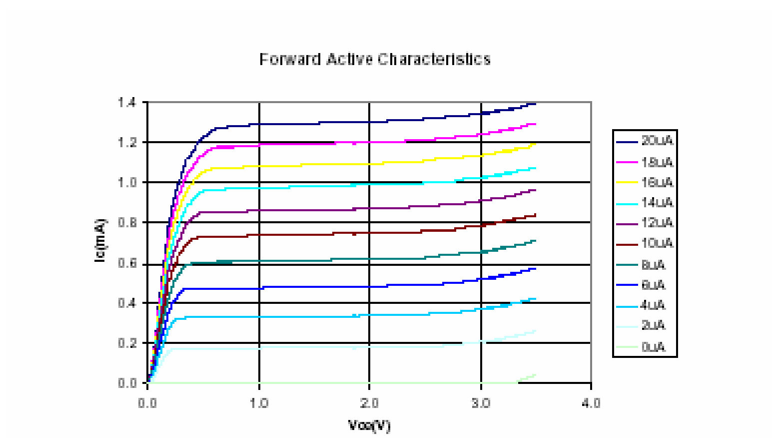

Figure 4.3: Forward active characteristics of the 2.5μm npn HBT at 25C.

Forward active curves for a typical npn HBT are shown in Figure 4.3. The early voltage, (V_A>60V), gives nearly flat curves from 0.5V to 2.8V of V_{CE}.

An important device parameter for very high-speed digital circuit operation is f_T – defined as the frequency at which the transistor exhibits unity current gain. It is related to the forward transit time at high frequencies. This value is a fairly strong function of bias current, and in order to make ensure maximum performance of the circuits, the bias current needs to coincide with the peak f_T.

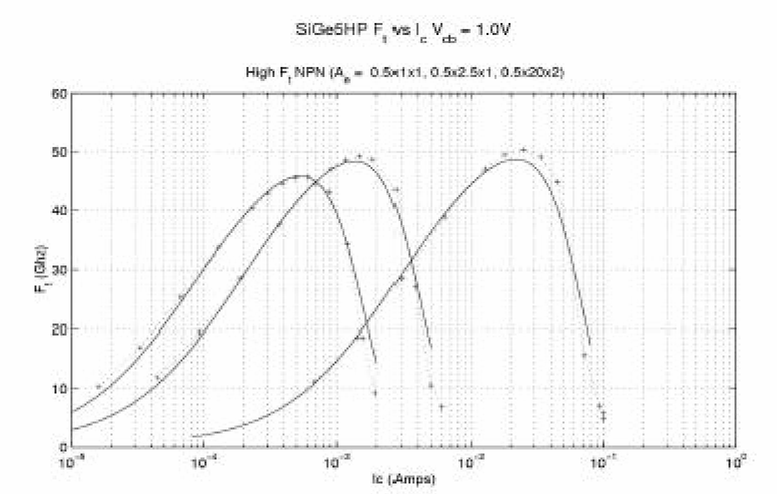

Figure 4.4: Plots from IBM showing simulated f_T as a function of collector current bias for three different sized HBTs at 25C.

As can be seen in Figure 4.4, larger transistors have a larger f_t, but also sink a larger amount of current. It is also important in that larger transistors will also induce a larger load on the circuits driving them due to the larger i_b, as well as have larger parasitic capacitance effects. From the above chart an approximate maximum f_T value occurs around I_C=0.6mA/µm for all transistors. Referring back to Figure 4.2, we can see that in the desired IC operating range, the forward VBE drop is between 0.85 and 0.9 volts.

Another important parameter is f_{max}, the maximum oscillation frequency. This is defined as the unity power gain frequency. It is related to f_T through the relation:

$$f_{\max}{}^2 = f_T / 8\pi R_b C_c$$

In the above equation, R_b is the base resistance and C_c is the base-collector junction capacitance. This parameter determines maximum bandwidth, which is important for the analog sections of the designs. According to the documentation, the peak f_T for the minimum sized device is 47GHz, while f_{max} is 65GHz at 75C.

Resistors

IBM's technology offers a variety of resistors with different sheet resistivities which range from 8.1Ω/☐ to 1750Ω/☐. Some of the other characteristics considered in choosing a type of resistor are the temperature coefficient(s), the parasitic capacitances, and the tolerances. The calculation of the actual tolerance for a particular resistor is quite complicated as the resistor instance is dependent on multiple mask and material tolerances, and the total resistance is dependent on a complicated geometry consisting of the primary resistive material along with contacts, vias, and other initial wiring. The polysilicon over deep trench resistor, (pbdtres), was chosen for general use because of it's relative temperature insensitivity, low parasitic capacitance, and good (relative) tolerance. It features 220Ω/☐, an almost zero temperature coefficient, and a low capacitance per unit area of 0.0667fF/μm^2.

In resistor layout, the length over width ratio determines the actual resistance. Ideally the resistor layout could be made as small as desired, but process variations will cause a relatively larger effect on a smaller physical resistor layout than on a larger one. When a larger resistivity material is used, this effect is increased. A spreadsheet was used to calculate tolerances of pbdtres resistors for particular resistance ranges used in the designs. A minimum layout dimension of 4µm for both the width and length of the polysilicon was found to be adequate to achieve tolerances on the order of 10%, with little incremental improvement beyond that. In the actual layout tool, one can specify a desired total resistance and the polysilicon width. To ensure a proper resistor, the width parameter should be increased until the calculated length is 4µm or greater. The current passing through the resistor must also be considered when choosing the geometry as the maximum current density is specified as 0.6mA/µm. As this is equal to the maximum f_T current parameter ratio for the HBT, the resistor width for circuits scales linearly with the transistor width for those circuits with transistors 4µm or greater.

Interconnect

IBM fabricates 5HP with several metalization options. Up to five metal layers may be used and there is an option, (5AM), available for creating a thicker last metal and insulating layer which is useful for creating higher quality inductors and low resistance wires. The 5AM option replaces the last metal layer(LM) and the one below that(M2,M3, or M4) with both the MT and AM(topmost) layers. A recently introduced option is 5DM, which is quite similar to 5AM except that the layer below AM is copper. The kit documentation for each option gives complete design criteria for creating reliable layouts. For a wire with a particular width and at a specific level the maximum safe DC and RMS currents can be calculated using the equations therein. Vias are also similarly covered. Most of the wiring in our designs falls into two broad categories, high current, and signal. The high current lines occur primarily inside gates and usually either carry a constant current or switch on and off with currents of at least 0.6mA. In some cases the currents go as high as tens of milliamps.

Table 4.1: TCR is %/C, (percentage change per degree C.)

Layer	Metal Thickness	Oxide Below	Min. Width	Max(mA)@ MinWidth	Max mA/μm	(Ω/□)	TCR %/C
M1	0.63±0.06	1.90±0.2	0.8	0.39	0.49	.076	.37
M2,M3,M4	0.85±0.08	1.2±0.31	0.9	0.95	1.1	.045	.39
LM	2.07±0.2	1.2±0.31	2.4	8.17	3.4	.015	.42
*MT	0.83±0.08	1.2±0.31	0.9	0.95	1.1	.045	.35
*AM	4.0±0.4	3.0±0.5	4.0	23.4	5.9	.00725	.38

*For 5AM option only. All distance units are given in μm.

In the table above, TCR is %/C, (percentage change per degree C). The values above are at 25C. Vias are 0.9μm square, and generally can handle 1.2mA each. For large current situations, a rectangular "via bar" is allowed. The selection of the 5AM option greatly eases the layout of power rails as they can take up one third the size if placed on AM as opposed to LM.

Other Components

The technology offers a large number of additional devices. A gated lateral PNP(GLPNP) transistor can be used. There are two types of capacitors, metal-insulator-metal(MIM), and decoupling. The MIM has fewer parasitics and is used whenever a capacitor is needed other than for power supply decoupling. The basic capacitance per unit area is 0.7fF/μm^2, with a maximum value as high as 7pF. Inductors are available in the 0.6nH to 15.8nH range with Q-factors as high as 10, or between 2.8nH and 83nH with Q-factors as high as 19 if the 5AM metalization option is chosen. The process also has Schottky barrier diodes(SBD) available with a 310mV drop and breakdown voltage of 5.5V. A P-i-N diode with 15V breakdown is available. In addition, a varactor diode with 1.37fF/μm^2 at 0V can be used. The varactor has a minimum available size of 6μm^2.

Logic

The digital circuits are implemented using differential current mode logic, (CML). This is essentially the same as differential ECL, but with the optional omission of emitter followers when feasible[15]. Analysis of this logic family draws upon [16], but is adapted for differential circuits as opposed to single-ended as is presented there. In a differential logic system, a pair of wires is used to indicate a boolean condition with the voltage difference between the pair indicating the state, (as opposed to the absolute voltage on a single wire.) The difference voltage, (V_{DIFF}), and common mode voltage, (V_{CM}), for a pair of wires (a_0,a_1) is shown below. Ideally, a differential logic system should be immune to variations in VCM and the outputs should be solely a function of VDIFF.

$V_{DIFF} = a_0 - a_1$
$V_{DIFF} = (a_0 + a_1)/2$

The individual voltage signals can also be expressed in terms of the difference and common mode voltages as:

$a_0 = V_{COM} + V_{DIFF}/2$

$a_1 = V_{COM} - V_{DIFF}/2$

Since all inputs and outputs are differential, there is no need for inverters since wires of a differential pair can simply be crossed to produce a complementary signal. The circuits mainly consist of current steering logic trees that use differential transistor pairs to divert current through one or another "branch" as it travels from V_{cc} to V_{ee}. Resistors at the "top" of the tree are used to generate voltage outputs that change when the current through a particular resistor's branch changes. These outputs swing through a voltage V_S, and drive other gates.

CML Gates

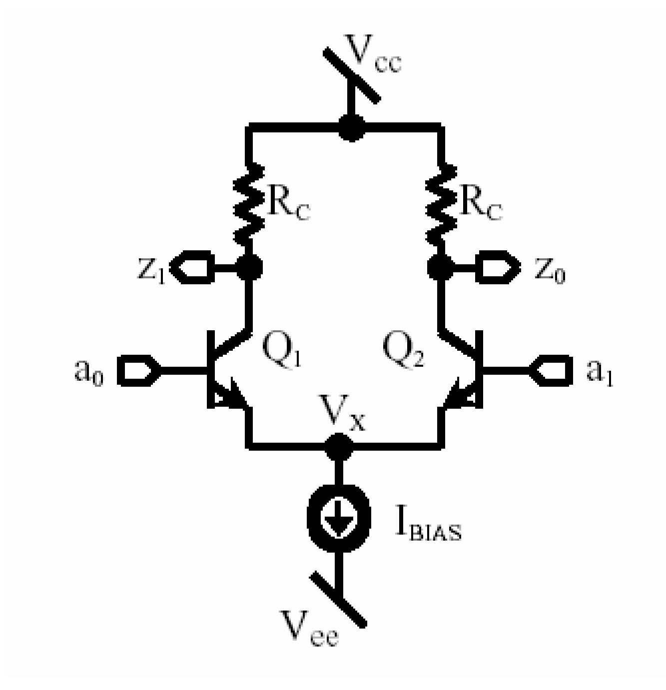

Figure 4.5: Simple CML buffer/inverter.

In Figure 4.5 is a simple buffer/inverter element. The current source at the bottom ensures that a fixed amount of current IBIAS flows through the tree at all times. In general, circuits that have different sized switching transistors will use a different value for I_{BIAS}, and the resistors at the top are chosen to provide a single design-wide uniform swing. The V_{CC} supply voltage is traditionally set at 0V, (which is the easiest to provide a stable value for), due to the direct effect this supply has on the signal outputs. V_{EE} is set to some negative value dependent on considerations that will be covered later.

A desirable property of current steering logic is that the constant current per gate prevents or minimizes the problems associated with switching noise on the power supply lines. This leads to smaller allowable logic swings, and ultimately to a smaller power supply voltage. The two complementary differential inputs, a_0 and a_1, cause the current to flow in one of the two branches through the collector resistors R_c, affecting the corresponding outputs z_0 and z_1.

The transistors in these CML trees are meant to remain exclusively in the active mode, so simplifications can be made regarding modeling their collector currents. In the standard Ebers-Moll model, with α approximately equal to one and the reverse saturation current I_{CS} negligibly small, we can express the collector current as:

$$I_c = I_s\left(e^{V_{BE}/\phi T} - 1\right)$$

Note that the I_S above is actually I_{ES}. If we were to solve this for V_{BE}, it would be obvious that the "1" term is insignificant since collector current I_C is much greater than I_S. Accordingly we can say:

$$I_c = I_s e^{V_{BE}/\phi T}$$

Using this, the ratio of the collector currents in the two branches are exponentially related to the difference voltage as shown in the equation below. The common mode voltage V_{CM} isn't a factor.

$$\frac{I_{C1}}{I_{C2}} = \frac{I_S e^{(a_0 - V_x}\!\!/_{\phi_T}}{I_S e^{(a_1 - Vx)}\!\!/_{\phi_T}} = e^{V_{DIFF}/\phi_T}$$

Figure 4.6 shows the percentage diversion as a function of the input difference voltage. A nearly complete swing is desirable due to the consideration that we wish to preserve and enhance the digital characteristics of signals. If we arbitrarily define "nearly complete" to mean 99%, the resulting current ratio in the equation above can be solved to give a minimum V_{DIFF} of only approximately 120mV, which is only about 5 thermal voltages at 25C.

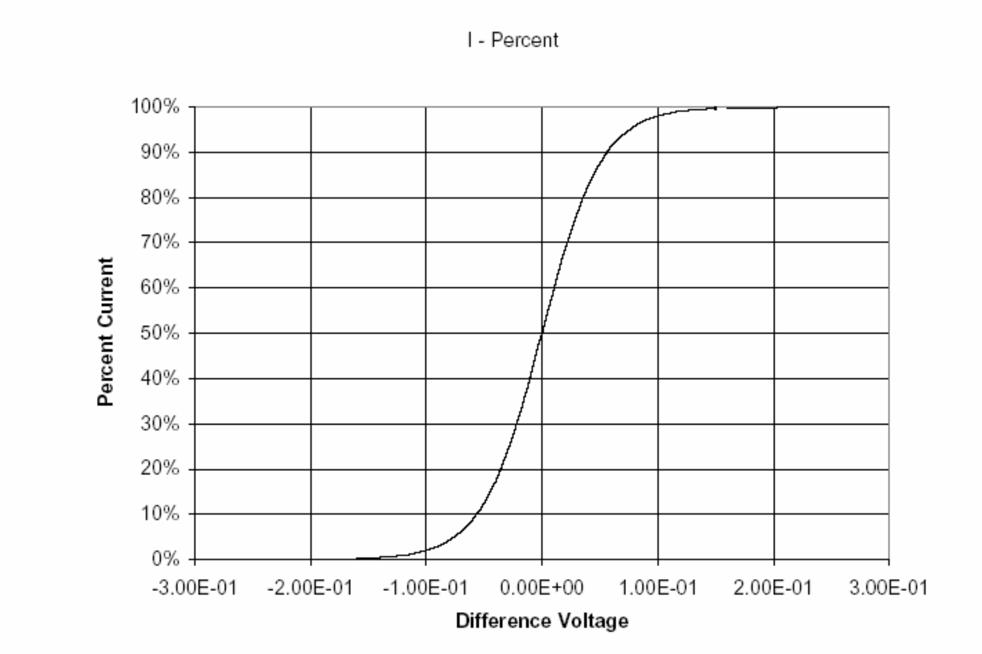

Figure 4.6: Plot of percentage of current down a branch as a function of the input difference voltage.

There is a dependence on temperature such that increasing temperature reduces the percentage of current diverted. At room temperature, (25C), a 60mV difference voltage is required to divert 90% of the current. At 155C, (the highest qualified temperature for hte SPICE models), nearly 100mV is required for the same effect.

$$I_{C1} = I_{BIAS} \frac{e^{V_{DIFF} / \phi_T}}{1 + e^{V_{DIFF} / \phi_T}}$$

The current ratio can be easily recast to obtain the single sided current as a function of the input difference as shown above.

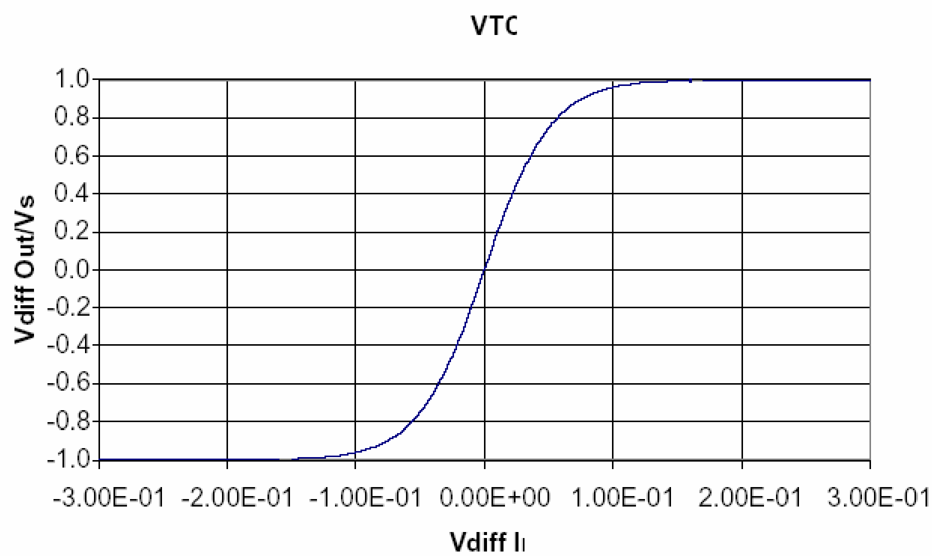

Figure 4.7: Normalized voltage transfer characteristic relative to chosen V_S.

Obviously when the full current is switched down one particular branch, the corresponding z_x output would switch between V_{cc} and V_{cc}-R_c* I_{BIAS}, which is referred to as V_S. The choice of desired output swing VS determines the RC resistor size. Figure 4.7 is a normalized voltage transfer characteristic(VTC). The gain in the middle region is about 18*V_S.

If the input voltage swing were not large enough, small changes in the input differential magnitude would cause larger changes in the output, amplifying noise. To prevent this, minimum high and maximum low input values(VIL,VIH) are calculated to determine when the gain is unity. (dVout/dVin=1)

$$V_{slope=1} = \phi_T \ln\left(\frac{-K \pm \sqrt{K^2 - 4}}{2}\right)$$

where

$$K = 2 - \frac{2R_c I_{BIAS}}{\phi_T}$$

Using the above equation, VIL and VIH can be determined. The differences between VIL and VOL(-VS), and between VIH and VOH(0V) are the static DC noise margins NML and NMH respectively[17]. They represent the magnitude of noise that could be present on a nominal output signal between a pair of gates that would still be valid at the input of the second gate. For a differential circuit, this would represent a change in the magnitude of the differential signal. It should be noted that the NM=0 boundary sets an absolute lower boundary of V_S at $2*\phi_T$. This is half of that calculated in [16] for single ended CML as expected. The low and high margins are equal due to the symmetrical nature of the circuits. A large noise margin is of course desirable, and this noise margin value grows with increasing V_S. However, increasing VS beyond a certain point would eventually result in saturation of the transistors at the top of the tree, resulting in a slower circuit. The transistors would enter saturation when $V_S=V_{BE}(on)- V_{CE}(sat)$. Providing a noise margin for that condition would require $V_S=V_{BE}(on)- V_{CE}(sat)-NM(sat)$. Solving for the noise margins simultaneously and plotting the results we obtain the figure below.

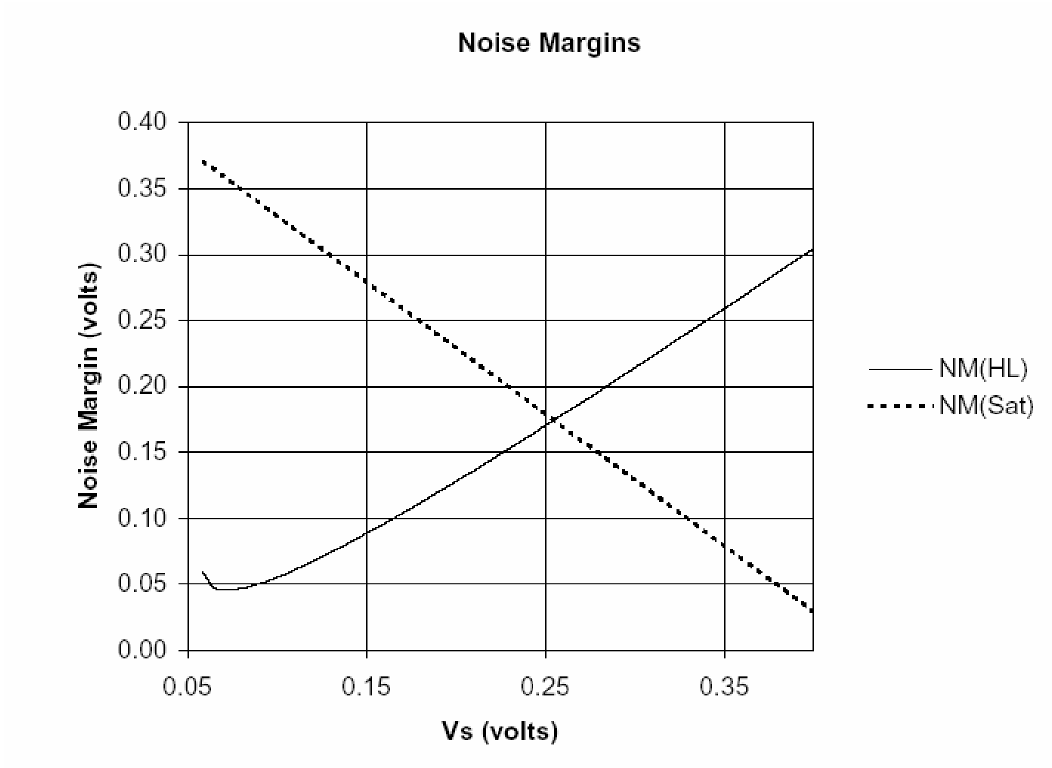

Figure 4.8: Optimal noise margin plot for determining V_S. The plot was derived using a temperature of 70C, V_{CE}(sat)=0.45V, and V_{BE}(on)=0.88V.

The intersection of the curves in Figure 4.8 indicates the optimal choice for V_S using the specified transistor parameters and temperature. The selection of 250mV for V_S seemed a good engineering choice as the optimal value varied between ~200mV and ~300mV as the temperature, V_{BE}(on), and V_{CE}(sat) were varied within reasonable ranges.

The choice of I_{BIAS} is determined by the f_T of the two identical transistors used in the circuit. Given the f_T curve in Figure 4.4, we choose to have approximately 0.6mA/µm based on the transistor size. The smallest size transistors would require R_C=416Ω resistors. Trees with 4µm emitter transistors would require an RC of 104 Ω.

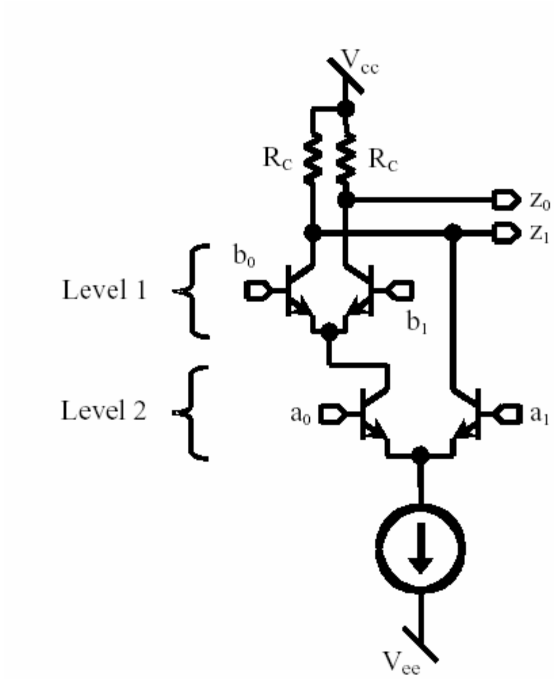

Figure 4.9: CML gate with two inputs.

In Figure 4.9, a 2-input, two level gate is shown. When the a_0 input is high relative to a_1, the current is steered by the b_x inputs and the z_x outputs change accordingly. When the a_1 input is high relative to a_0, the z_1 output is pulled low regardless of the values at the b_x inputs while the z_0 goes high. This single circuit can be used as a logical AND, NAND, OR, or NOR, merely by labeling the pins accordingly. Two level trees can also be constructed to form any two input combinatorial logic function and can even incorporate feedback to form latches as in Figure 4.10.

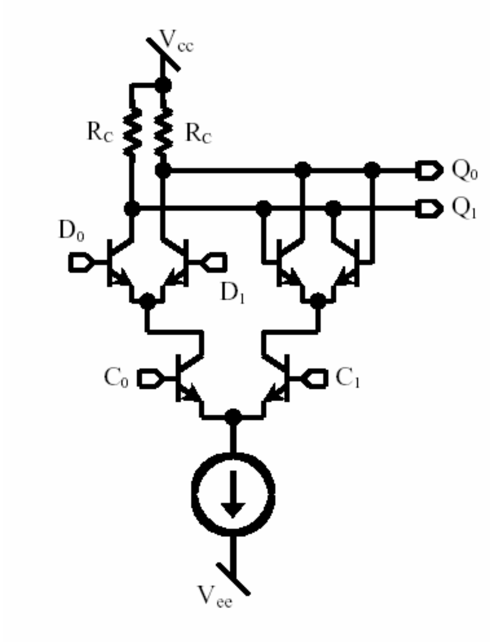

Figure 4.10: CML D-Latch

When C_0 is high and C_1 is low current is steered through the left level-one transistor pair. Inputs D_0 and D_1 are then active and cause a corresponding output at the Q_0 and Q_1 terminals. When C_1 is high and C_0 is low, the current is steered through the right upper level pair of transistors and these feedback transistors hold the previous state of the D_X inputs. Because there is no current in their branch the D_X inputs are no longer active. This stacking of differential pairs is called "series gating," and it is another reason to require nearly complete switching of current. The percentage of current switched is repeated at each level, and if only 90% of the current were switched at each level in a three level circuit, less than 75% would be available at the top to generate the output swings. With 99% switching we have 97% of IBIAS available at the top.

In general, an N-level tree can have up to $2^{(N-1)}$ inputs and can be used to create any N-input combinatorial logic function. We use up to three levels in the designs. One side effect of adding multiple levels is that the propagation delay via lower level inputs is generally longer than that for upper ones due to the loading of the transistors above them. Another unfortunate effect of adding levels is that the signals to a set of inputs effectively sets the collector voltage for the pair of input transistors below them. If the lower level inputs had the same common mode voltage as the upper ones, their transistors would saturate with VCE□0V, and the circuit switching performance would be impaired.

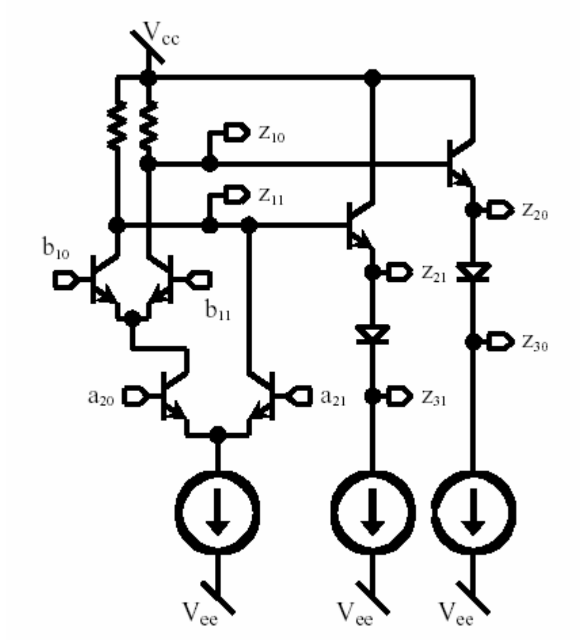

Figure 4.11: Two level gate, with emitter followers providing different output levels.

To address this issue, when lower level common mode voltages are needed, emitter followers like those in Figure 4.11 are added. The level naming convention is also illustrated, with inputs and outputs having subscripts indicating the level (1-3) and whether it is the main signal line(0) or it's complement(1). Each junction in the emitter follower drops the output signal by VBE(on), preparing it for driving gate inputs which require that level. . The emitter followers also improve the single-sided rise times over the passive resistor pull-ups. In order to adequately drive different loads, the transistors in the emitter followers can be of a different size than those in the main tree. The bias current for these followers would also be adjusted accordingly.

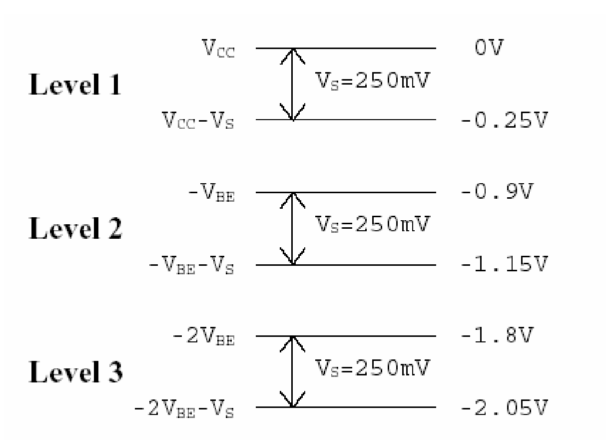

Figure 4.12: Nominal voltage levels for CML inputs.

The chosen voltage levels for the three level circuits is illustrated in Figure 4.12. The large separation between levels relative to the voltage swing ensures common mode noise will not induce saturation for the transistors on the level below. It should be noted that the inputs to gates with less than three input levels need not necessarily use the specific levels designated, but rather as long as the levels are separated and in the proper order the circuit will work correctly. The levels differ only by a common mode value, and CML is generally indifferent to that. Because of this, a single level buffer can accept a Level 1, 2, or 3 input signal. A gate with two input levels can accept the level pairs: 1 & 2, 2 & 3, or even 1 & 3. In circuits with 4 levels there would need to be additional restrictions due to the possibility of exceeding the collector/emitter breakdown voltages. A Level 4 signal into a simple single level buffer would cause V_{CE} on the input transistor to be 3.34V, which is above the minimum rated breakdown voltage.

It should be noted that fan out for these circuits is very large, in the sense of static driving capability. The emitter followers can be increased in size to handle almost any desirable number of loads. With large numbers of loads though, the rise, fall, and propagation delay characteristics suffer. Since the primary impetus for using CML is for performance, we limit loading for mostly speed considerations, and size the transistors accordingly.

Current Sources

One final piece for the logic family is required, and that is the current sources positioned at the bottom of each of the trees. It was decided to use an active current pull-down as opposed to a passive resistor for performance reasons. Simulations show a definite performance increase, especially under high loading conditions.

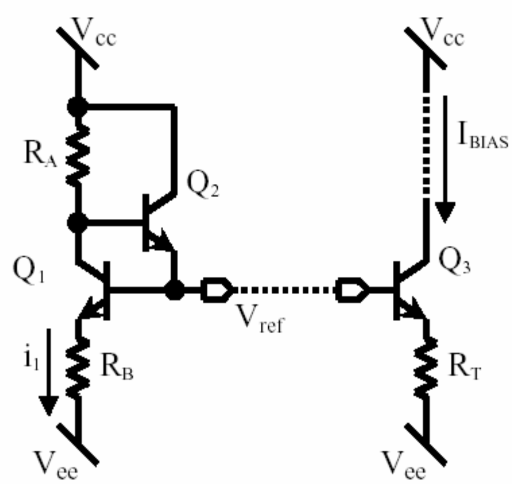

Figure 4.13: Current Mirror

The constant current source at the bottom of the tree is provided using a Widlar current mirror circuit. In Figure 4.13, the circuitry on the left provides a nearly fixed voltage reference, V_{ref}, by using forward biased junctions and resistive drops. On the right, the IBIAS current line represents a gate or emitter follower circuit as described previously. In some implementations of a current mirror, Q_2 is absent and a base-collector short is placed across Q_1, effectively turning it into a diode. The presence of Q_2 allows greater current drive capability, with nearly all of the V_{ref} drive current arriving via its collector. The V_{ref} voltage is $i_1 * R_B + V_{BE}(on)$ above the V_{EE} rail, and is derived via the following equation:

$$V_{ref} = V_{EE} + V_{BE}(on) + R_B * \frac{V_{CC} - V_{EE} - 2 * V_{BE}(on)}{R_A + R_B}$$

Since the reference isn't switching, the current i_1 is sometimes set to a value below that of the current maximizing f_T in order to save power. The tail resistor R_T can to adjust the current for individual trees serviced by the same $V_{ref.}$ This is useful as trees using larger transistors require a larger bias current in order to maximize their performance. For a particular tree,

$$I_{BIAS} = \frac{V_{ref} - V_{BE}(on) - V_{EE}}{R_T}$$

The derivative of the above with respect to V_{ref} shows the need to have a larger R_T if we want to reduce the variations in the bias due to variations in the reference. One of the V_{ref} generators can supply a reference voltage to many gates because of the small base currents involved. If the loads is given in terms of the total microns of emitter length, the load microns can have a 50:1 ratio over the reference microns if the peak f_T current is used in the reference, neglecting things like voltage drops due to wiring.

The collector voltage at Q_3 can be fixed as low as -2.7V by a level three signal above it. Avoiding saturation in Q_3 effectively sets a maximum allowable value of V_{ref} as –2.25V if V_{CE}(sat) is considered to be 0.45V and V_{BE}(on)=0.9V. The emitter of Q_3 is then one V_{BE} less than V_{ref}, or –3.2V. Choosing the difference between this value and V_{EE}, (deciding on V_{EE}), determines the size of the individual R_T resistors. This choice is somewhat arbitrary, but the size of the tail resistors should be on the order of that of the RC resistors in the current trees, as the same tolerance concerns apply here as there. The trees with the smallest transistors will require the largest resistors, and it was decided that the voltage developed across the tail resistors for these should be roughly the same as the voltage swing. Thus, V_{EE} was decided to be equal to –3.2V-V_S, or 3.45V. As V_{EE} is lowered, the chip power requirements grow proportionally as the chip is designed to use a constant current. In the actual design of the SERDES II chip, the V_{EE} rail was designed to be -3.4V, with the V_{ref} voltage designed to be around -2.2V. The V_{CE}(sat) value of 0.45V is conservative, and the current source transistors do not in fact become saturated. Additionally, the highest speed circuits tend to use less than three levels. In retrospect, and in future designs, plan to design for a -3.6V or higher V_{EE}, to allow a common mode noise margin for the Level 3 inputs.

Another design concern is the maximum V_{CE} which can appear across the current source transistor at the bottom of the tree, which is –V_{BE}(on)-V_{EE}-VS=2.45V for V_{EE}=-3.6, or 2.25V for

V_{EE}=-3.4V. The first SERDES design was planned around a –4.5 V_{EE}, which would have resulted in a maximum V_{CE} of 3.35V, which would have exceeded the breakdown voltage of the normal HBT(npn). For that reason, the SERDES and SERDES II designs used a high breakdown HBT(npnhb) as the current source, as this transistor has a minimum breakdown of 4.5V. The f_T of this transistor is drastically reduced however, and it is thus a less responsive current source. Also, the bias current for the maximum f_T for the npnhb is only 1/6th that of the npn. To achieve optimal performance the npnhb would need to be six times larger which would increase parasitic effects and layout requirements unreasonably. For these reasons, future designs will use a regular npn as the current source.

Performance

The performance characteristics of the CML logic family using this technology are extreme. Unloaded gate delays on the order of 10-20ps are certainly achievable, leading to clocking rates as high as 10GHz. Due to the excellent device performance, the circuits can be limited by the parasitic effects of interconnect almost as much as by device characteristics. Below are plots showing simulated simple buffer performance both with and without the RC parasitic effects from actual layout interconnection, and how performance is affected by the addition of emitter followers. The simulations were all done at 75C, and used the built-in parasitic extraction engine in Cadence which creates a RC ladder network for each interconnection.

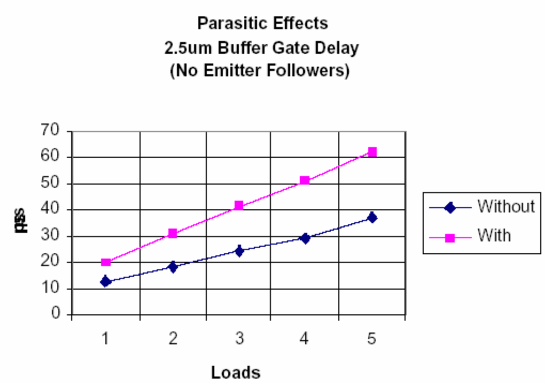

Figure 4.14: Plot showing large effect of interconnect on buffer performance for differing numbers of loads.

As can be seen in Figure 4.14, the interconnect can degrade performance by as much as 33%. In practice, the amount of degradation is closer to 10%-15% as the simulation layouts used in these plots were not as tightly laid out as are the circuits which were fabricated. The interconnect effects are largest for the smaller transistors(1μm) of course, but increasing the transistor size tends to yield marginal improvements on tight layouts when the transistor sizes reach around 6μm.

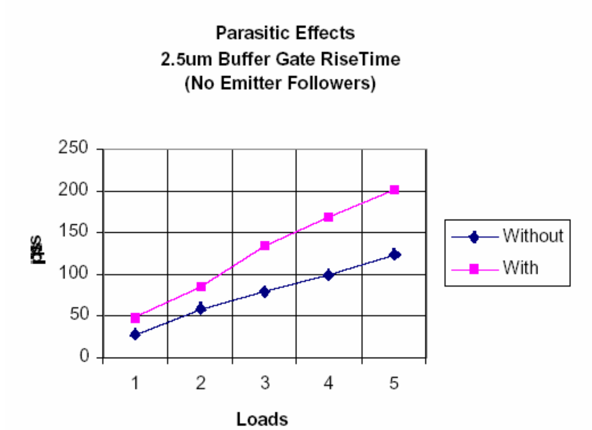

Figure 4.15: Interconnect effects on buffer rise/fall times for various loads.

In Figure 4.15, it can be seen that the interconnect has an even larger effect on the rise and fall times of gates. Luckily, the introduction of emitter followers greatly enhances the driving capability of the circuits. This is shown in the next figure.

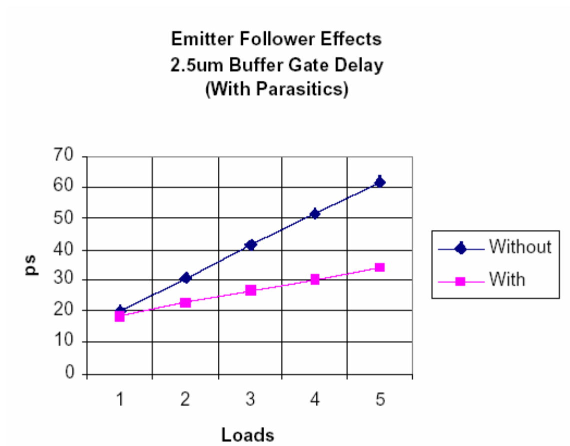

Figure 4.16: Gate delay reduction in 2.5μm buffers with the use of 2.5μm emitter followers.

As can be seen in Figure 4.16, the introduction of emitter followers greatly reduces the effective delay in CML gates, for nearly all conditions. In fact, the only time it has been observed to actually increase the effective gate delay has been when only a single load is used, and the interconnect parasitics were small. Under those circumstances the added delay of going through the emitter follower transistors was more than that made up for by the extra driving capability.

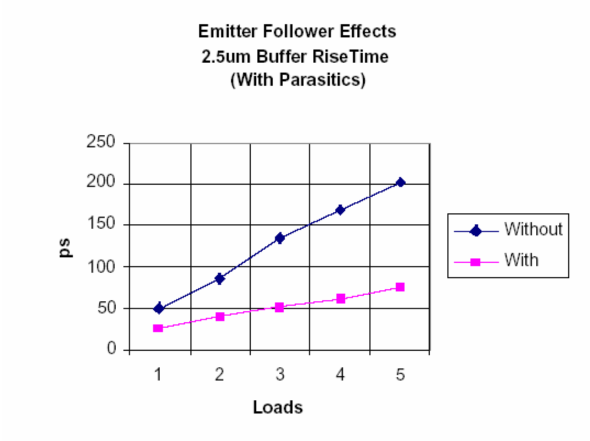

Figure 4.17: Effect on rise time of adding 2.5μm emitter followers to 2.5μm buffers.

With emitter followers added, the rise and fall times are greatly improved as well. The only real disadvantage to the use of emitter followers is that the signals must always output on level 2 or lower, and the emitter follower circuits often more than triple the power consumption of a given gate. However, since the focus is on raw performance, we use them quite liberally in the designs. There is an increase in layout size as well, but as we are not creating large circuits it is not of primary concern at this time. As mentioned in the logic section, the emitter follower transistors can be of a different size than the transistors that make up the logical functioning part of the gate, and should be biased accordingly. Increasing the size of the emitter followers improves the driving capability of the gate, but with diminishing returns. In addition, larger emitter followers have a greater load burden on the logic transistors, and beyond a certain point this will actually degrade performance.

High Frequency Response

Another concern in the design of the circuits is the effective bandwidth of buffers, gates, and multiplexers. Because we are operating close to the effective limits of the technology, we need to consider that the circuits might not return to a steady state condition in a single bit time. This is the direct result of bandwidth limitations, and the effects are expressed as data-dependent jitter.

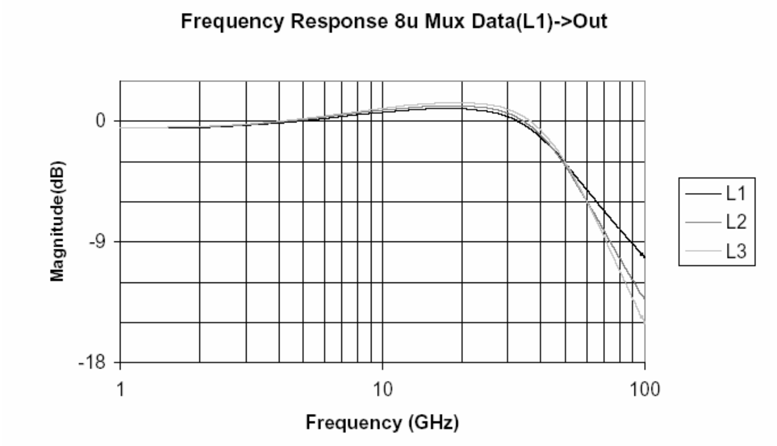

Figure 4.18: Large signal bandwidth of 4μm buffer with three output levels.

Note that the normal small-signal bandwidth calculation methods will not result in an accurate representation as the circuits do not operate in a small signal fashion. The plot in Figure 4.18 was generated by sweeping the input frequency of a sinusoid in the time domain. The amplitude and phase were extracted using an ahdl, (analog hardware definition language), program. The simulation was performed on the data inputs of a 2:1 MUX, as opposed to the select input, which generally has a smaller bandwidth.

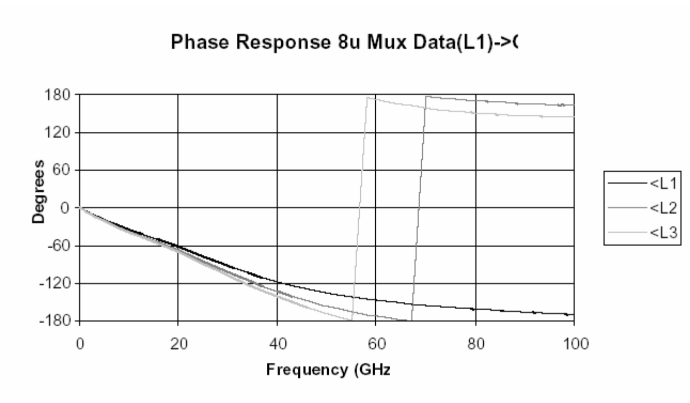

Figure 4.19: CML buffer phase response plot using transient large signal swing simulation.

Phase plots are important for determining sources of pulse distortion. As can be seen in Figure 4.19 the phase response of the buffers grows more non-linear at the higher frequency extremes. The L1 output signals show the greatest distortion, due to their inferior current sourcing capability.

Linearized Differential Buffers

In some circumstances it is necessary to either reduce or eliminate the gain from a common differential pair, for example, when attempting to increase bandwidth. This can be accomplished by the addition of emitter resistors, RE, as shown in the figure below.

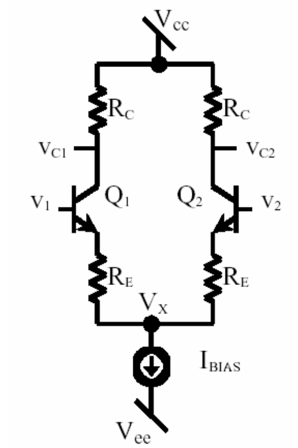

Figure 4.20: Linearized buffer circuit with emitter degenerating resistors.

The differential transfer characteristic can be derived from the loop of voltage drops encompassing the base-emitter junctions, the emitter resistors, and the input difference voltage $v_{ind}=v_1-v_2$. Assuming the following:

$$I_E = I_{ES}e^{V_{BE}/\phi_T}$$

V_{BE} can be cast in terms of the emitter current, and the loop equation can be solved for v_{ind} to give:

$$v_{ind} = R_E i_{outd} + \Phi_T \ln\left[\frac{I_{BIAS} + i_{outd}}{I_{BIAS} - i_{outd}}\right]$$

$$v_{outd} = -aR_C i_{outd}$$

$$v_{outd} = v_{C1} - v_{C2}$$

$$i_{outd} = i_{E1} - i_{E2}$$

$$v_{ind} = v_{be1} - v_{be2}$$

In the equations above, i_{outd} is the difference current i_1-i_2. Multiplying this difference current by the collector-emitter current ratio alpha, and by the value of the collector resistors R_C, the input value v_{outd} can be plotted in terms of the output value v_{ind}. With the axis exchanged, the familiar transfer characteristic can be displayed.

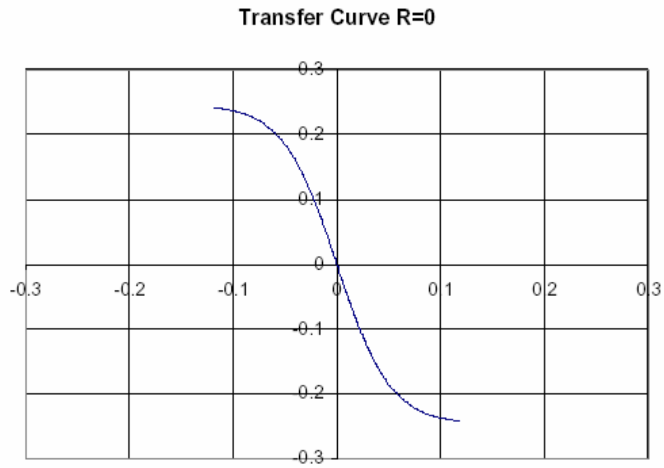

Figure 4.21: Inverter transfer characteristic without emitter resistors.

The figure above represents the transfer characteristic of a non-linearized one micron buffer biased at 0.6mA using T=25C. The collector resistors are 416Ω and β=86 yielding α=0.99. This base plot with no emitter resistors can be compared with the figures below. Note that the derivative of the previous equation can be taken and it yields the reciprocal current gain, which we can recast into the buffer voltage gain.

$$\frac{dv_{ind}}{di_{outd}} = R_E + 2\Phi_T \frac{I_{BIAS}}{I_{BIAS^2} - i_{outd^2}}$$

$$i_{outd} = -\frac{v_{outd}}{\alpha R_C} \approx -\frac{v_{outd}}{R_C}$$

$$\frac{dv_{outd}}{dv_{ind}} = \left[\frac{dv_{ind}}{di_{outd}}\right]^{-1} \left[\frac{di_{outd}}{dv_{outd}}\right]^{-1} = \left[\frac{1}{R_E + 2\Phi_T \dfrac{I_{BIAS}}{I_{BIAS}{}^2 - \left(v_{outd} / \left(\alpha R_C\right)\right)^2}}\right] \left[-\alpha R_C\right]$$

At the midpoint of the transfer characteristic, this reduces to:

$$\left.\frac{dv_{outd}}{dv_{ind}}\right|_{v_{outd}=0} = \frac{-R_C}{R_E + \dfrac{2\Phi_T}{I_{BIAS}}} = \frac{-R_C}{R_E + 2/g_m}$$

This gives a gain of 4.8 for the 1 micron non-linearized buffer described above. Unity gain can be achieved with a 373Ω RE.

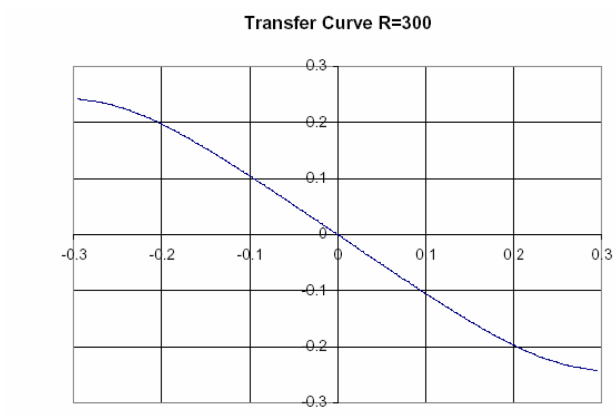

Figure 4.22: Linearized buffer transfer characteristic.

The same buffer with 300Ω emitter resistors is shown in Figure 4.22 above. Note the near linearity around the center. However, at the extremes, voltage gain can be less than unity. It is also of concern that the addition of emitter resistors will effectively reduce the voltage at the base of the tree V_x, just above the current reference. This voltage will also swing through a much larger range. If a current mirror is used to provide the bias, care must be taken to ensure that the transistor is not driven into saturation.

Pad Drivers

The chips have for the most part been probed on wafer, using 50Ω probes, sma cables and test equipment. Due to the limitations of the testing set up, we were limited to examining at most one high speed (>1GHz) differential signal at one time. We made use of DC input control signals to change on-chip configurations to present different output characteristics at the high speed pad. We also had a number of medium-speed outputs and inputs which we used for sending in reference signals for PLLs, or to obtain a trigger signal for generating an "eye" diagram of a high speed output.

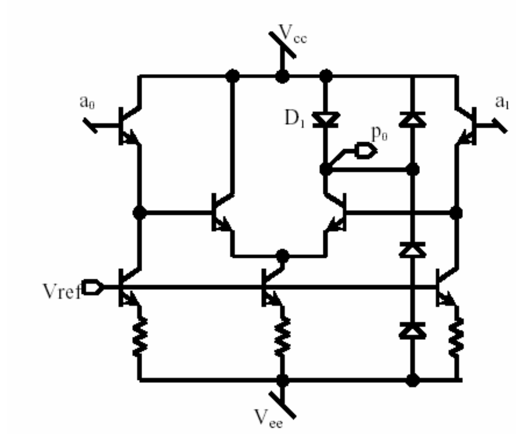

Figure 4.23: Single-ended medium frequency pad driver for signals under 1 GHz.

Above is a medium speed single ended pad driver. Not shown is a 1 μm buffer input stage with outputs to a_0 and a_1. A large voltage swing is developed across diode D_1, providing a signal sent to a bond pad via p_0. The other diodes are present for electrostatic protection, but regular diodes are used instead of the provided large ESD components in order to reduce the amount of parasitic capacitance introduced. The central core has 6μm transistors, while the emitter followers, (actually, leaders as shown), are 2μm. With $6*I_{BIAS}$ current across the diode, we obtained nearly 800mV DC swings. The pad itself is a large LM plate with a DT mesh or NS plane beneath, and it is modeled as a capacitance with a parasitic reverse biased diode to the substrate.

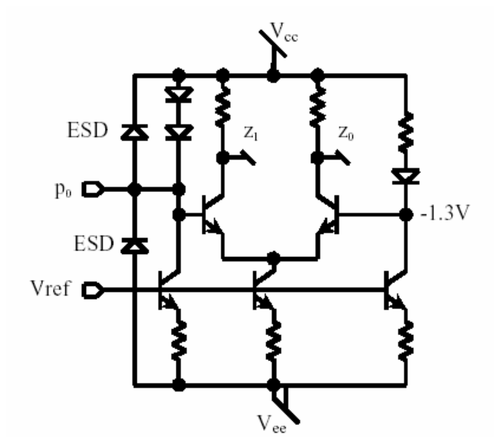

Figure 4.24: Single-ended DC signal pad receiver with ESD protection. This pad driver is designed to go to a default state if left floating.

Figure 4.24 shows a single ended DC control signal pad receiver. It uses the design kit suggested

ESD diodes for static protection. The central buffer uses 1µm, and the z_x outputs go to 1µm emitter followers (not shown). These emitter followers are small because they are intended to drive a single load at level two, and the input is not intended to change during normal circuit operation. The other diodes and sources are 2µm sized and provide biasing so that the current in the central buffer switches when the input is varied around –1.3V. When not connected, p_0 will be pulled to –1.7V, and z_1 will be brought low. The medium speed input circuits are similar, but have a 50ohm matching resistor added at the pad.

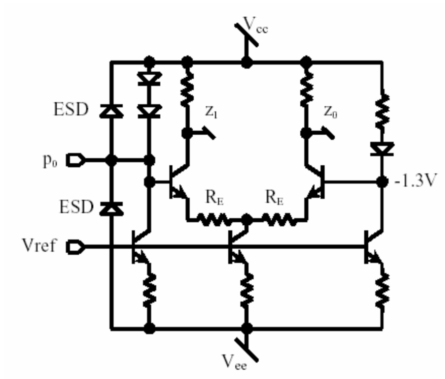

Figure 4.25: Linearized analog DC signal pad receiver. Allows linear input voltages to be passed differentially to internal circuits.

The circuit above is identical to the previous one, except that it incorporates 100 ohm emitter degenerating resistors R_E, which linearize the buffer response. This is used to directly apply a bias signal when desired, such as when a VCO is controlled by an external analog voltage.

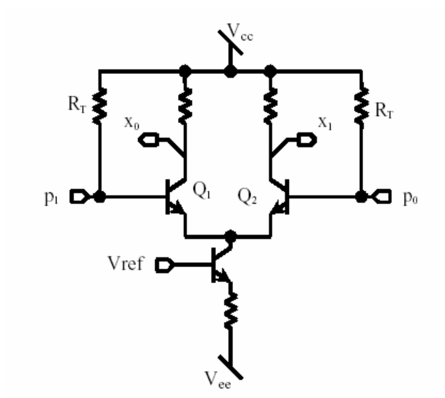

Figure 4.26: High-speed differential pad receiver circuit with resistive termination on chip.

Differential high speed pad receiver with RT=50Ω input resistors. It is a simple 6µm buffer with resistors sized accordingly, but with 50ohm resistors added for input matching.

In order to drive the signals off chip, we need to consider transmission line characteristics. The test stand used probes and cables with 50Ω characteristic impedances (Z_0). The test equipment is assumed to be matched, which means the effective load the driving circuits will encounter is equal to Z_0. Ideally, the source resistance(Z_S) of the output circuits should be 50Ω as well, in order to attenuate any signals reflected back into the output pads. As a first approximation for this, the collector pull up resistors should be 50Ω. This has the effect of creating a 25Ω(Two 50Ωin parallel), effective load for the collectors to drive. Developing a voltage of 400mV would mean supplying as much as 16mA. Because of this, we have opted in some cases to use high impedance output circuits and trust the end termination to hopefully absorb the signals. Future work would involve improving on this. At 40Gbps, the *fundamental* frequency is 20GHz. The signal is made up of frequency components many times higher. At 5 times the fundamental, 100GHz, the wavelength is ~1.5mm.

Power Rails

The design of power rails for logic cells arranged in long rows is straightforward if the power usage is fairly even over the length of the rail. This is true for the majority of the CML logic circuits. We can derive some equations useful in design. Consider Figure 4.27, which shows the relationship between the voltage along the V_{CC} power rail as a function of distance.

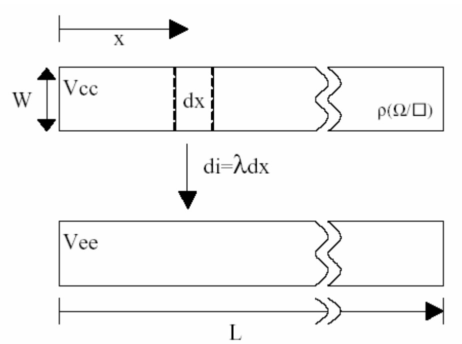

Figure 4.27: Power rail dimensions and characteristics.

Assuming the circuit is powered from the left, $V(0)=V_{cc}$. The parameter λ is a measure of how much current is drawn by circuitry along the rail. This can be arrived at by looking at the amount of current per average cell, divided by that cell's length. If you examine a long row of cells, calculate the total current and divide it by the length of the row. The parameter ρ is the sheet resistivity of the metal layer the power rails appear on.

$$\%droop - \frac{V_{CC} - V(L)}{V_{CC} - V_{EE}} x100\%$$

Droop is the percentage of the original voltage lost to resistance in the power rails. The above value should be set to some reasonable limit, such as 1%. It should be noted that this value is for one-sided supply droop only, the reduction in effective supply voltage is actually twice this. We find the droop by examining the differential element 'dx'. The current through that element and the material characteristics cause the change in voltage. The current varies along the rail as follows:

$$i(x) = I_{TOTAL} - \lambda x = L\lambda - x\lambda = \lambda(L - x)$$

Next, the voltage drop across the differential element is given as:

$$dV = i(x)\frac{\rho}{W}dx = \frac{\lambda\rho}{W}(L - x)dx$$

The integration of differential voltage changes, with the addition of the boundary constant V_{CC}, gives the voltage at any given point:

$$V(x) = V_{CC} - \frac{\lambda\rho}{W}\int_0^x (L - \mu)d\mu = V_{CC} - \frac{x\lambda\rho}{W}\left(L - \frac{x}{2}\right)$$

We are interested in V(L), which when evaluated and used in the previous droop equation gives:

$$\%droop = \frac{L^2\lambda\rho}{2W(V_{CC} - V_{EE})}x100\%$$

This equation allows a number of different possibilities to be explored. It can be solved for the maximum length, the necessary width, etc. For illustrative purposes, a high value for λ might be 87.5A/m. The rail width(W) might be 20μm. V_{CC}-V_{EE} might be 3.6V. Lastly, the sheet resistivity of the LM layer is 0.015Ω/☐. Using a 1% power droop criterion, we can find the maximum allowable length of the rails to be 1047μm. If the power were distributed on one of the intermediate metal layers, M2..M4, the maximum safe length value would drop to 605μm. Connecting the power rails at both ends would double this length, while requiring a minimum 1% change in total supply voltage would again halve it.

Chapter 5 Clocking Components

To realize complete SERDES systems, we need more components than just logic gates, buffers, and pad drivers. For the receiver, a local oscillator must generate regular pulses so that the input data signal can be sampled at regular intervals. This clock must adapt to variations in the input in such a way that bit errors are minimized. The transmitter requires an accurate clock to generate a clean data signal for the receiver. Because highly accurate oscillators are difficult to create on-chip, the internal transmitter clock is usually derived in some way from an accurate external reference. The adaptation of the local VCO output to an external source is handled by a phase locked loop circuit in both cases.

Phase Locked Loops(PLLs)

Phase locked loops form crucial elements of SERDES circuits. In simplest form they consist of three primary components; a phase detector(PD), loop filter(LF), and a VCO(Voltage Controlled Oscillator). These elements operate together in a closed loop control system as shown in Figure 5.1. The phase detector detects a difference in phase between the input and output signals and generates a phase error signal. This phase error is passed to the loop filter that filters out high frequency components, and has a high low frequency gain. The output of the filter becomes the VCO input signal, either increasing or decreasing the frequency until the signals are again in phase. Often a frequency divider is employed in the feedback loop so that the VCO generates an output that is some multiple of the input frequency.

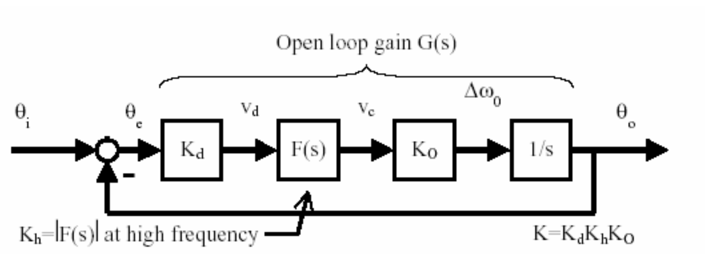

Figure 5.1: Simple phase locked loop diagram.

In control system terminology, this is a single-input, single-output closed loop system with the variables of interest being the phases of the signals. Since frequency is the derivative of the phase, holding the phase differences at a constant value causes the frequency differences to be zero as well, and the output frequency will track the input over some acceptable range of values. The model is usually linearized, and the VCO is modeled as an integrator. Standard control system analysis usually results in a second order model.

Phase Detectors

The phase detector is the circuit that compares the input signal to the signal from the VCO and produces an output signal conveying information regarding the phase difference between the two. An "ideal" dynamic detector circuit would produce a linear output over the range $\pm\pi$, as it is usually not possible to detect more than one full period of phase difference since the signals are assumed to be periodic. (Actually, you can have such a detector, but it must incorporate internal states.) There is typically a gain K_d associated with the phase detector that adds to the total open loop gain. In addition, when the phase detector has no input signal applied, it is often useful to have the output of the phase detector be the same as it is when the signal is exactly in phase, which is called the *free running voltage,* V_{do}. This prevents the absence of the input from immediately causing the VCO frequency to change. In situations like clock extraction, this is very important. The parameter V_{do} also in part determines the PLL's static phase error, and it is desirable to have this parameter be close to zero[18].

One common type of phase detector is simply a multiplier, whereby one signal is used to modulate the other producing a signal with spectral components centered at twice the input frequency and at a frequency of zero. The higher frequency spectral components are either naturally attenuated or intentionally filtered out, ideally leaving a DC signal proportional to the sine of the phase error, if the inputs were originally sinusoidal. For low values, the sine function is approximately linear. The indication of phase error polarity is correct over the range $\pm\pi/2$. A Gilbert Multiplier is an example of this type of phase detector. It suffers from the disadvantage that the magnitude of the output is in actuality dependent on the magnitude of the inputs.

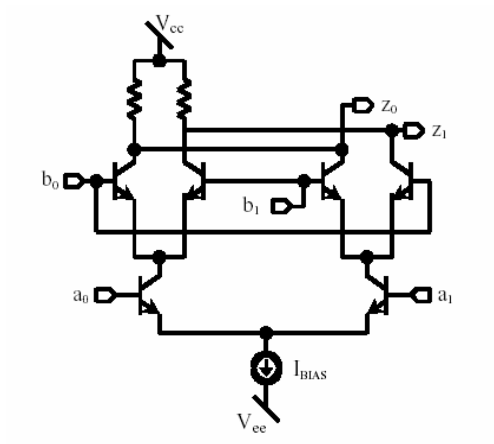

Figure 5.2: Gilbert Multiplier/XOR phase detector

A CML XOR gate has an identical circuit diagram when compared to the Gilbert Multiplier, but when used on digital signals it's outputs swing from one extreme to the other, removing the dependence on input amplitudes. This is referred to as an XOR phase detector. With square input pulses, the XOR phase detector produces a digital stream of pulses that have a duty cycle proportional to the difference in phase. This pulse signal has an average value that can be used as the detected phase and it is linear over a $\pm\pi/2$ region.

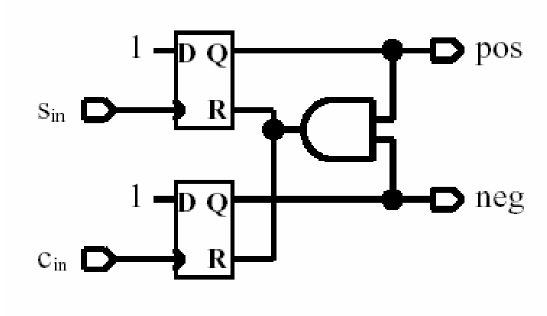

Figure 5.3: Three state phase detector. The two outputs indicate edge position and can be combined differentially.

Another common type is the 3-state phase detector, (Figure 5.3.) This is a digital circuit that incorporates latches, and thus can indicate a phase error greater than $\pm\pi$. It is simple and works well, which is the reason for its popularity. It incorporates two edge-triggered latches and gives output pulses with widths that are proportional to the time between the input pulse rising edges. The average of the output(s) is a linear signal valid over $\pm 2\pi$ radians. These outputs can be used as a single differential signal. This is a good phase detector for clock synthesis purposes, and it is what we use in the designs.

The above phase detectors are mostly suitable for clock synthesis, or for other uses when the PLL input is regular. For the receiver designs, we have to consider the case when edges in the bit stream are not present. When this occurs, the phase detector must supply the same signal as if the input edge arrived in perfect synchronization with the local clock edge. To achieve this, we over-sample the input stream by a factor of two, and compare the values of adjacent samples to determine if an edge is present. The sampling takes place in a round-robin fashion using eight sample latches. If the edge is present, a "fast" or "slow" pulse is generated. Since there are many data sample latches, there are many fast and slow lines. The fast and slow pulses are summed, and the resulting averaged signal is sent to the loop filter. This is in essence a "bang-bang" phase detector when a single edge is examined, but the averaging over several edges smoothes out the response.

Loop Filters

The loop filter is the primary design control point. By modifying this filter, the behavior of the PLL control system is adjusted. A resistive dividing attenuator can reduce the again, and hence the bandwidth, but the dc-gain is affected as well. This tends to reduce the signal or gain swing available to the VCO. A passive loop filter using resistors and capacitors can affect the ac and dc gains separately. However, a PLL loop filter is frequently implemented as a proportional-plus-integral active circuit, using an op-amp. By designing the filter to have infinite gain at DC, the static phase error can be made independent of the oscillator center frequency, (V_{CO}), which is a difficult parameter to control.

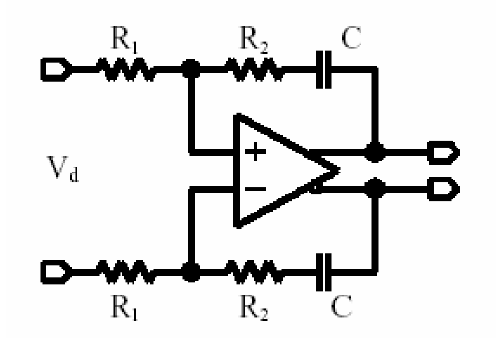

Figure 5.4: Differential active loop filter

The circuits use differential implementations of phase detectors, loop filters, for the VCO as well. In Figure 5.4 the basic circuit for the active differential filter is shown. This circuit has very high gain at low frequencies, and a flat gain K_h for high frequencies, given by the following equation.

$$F(s) = K_h \frac{\omega_2 + s}{s}$$

$$K_h = R_2 / R_1$$

$$\omega_2 = 1 / CR_2$$

The op-amp introduces an additional pole that eventually causes the output to decay, giving a low pass characteristic with a large pass-band. We actually introduce an earlier pole using a preceding RC low pass filtering stage so the op-amp pole doesn't come into effect. The op-amp is realized using an nfet differential buffer followed by a CML buffer with emitter followers.

The primary purpose of the loop filter is to adjust the PLL bandwidth. This changes how well the local VCO can track the incoming reference signal. PLLs can be tuned to screen out undesirable effects. For example, if the input reference clock has an accurate average frequency, but possesses more phase noise, (undesirable high frequency components), than the local oscillator, the loop bandwidth should be made small so that the output will have accurate frequency, with less noise. The loop bandwidth directly affects the step response, so the lock-in time of the PLL is affected.

VCOs

The designs make use of voltage controlled ring oscillators. Ring oscillators were chosen over other types because of the multiple phase requirements of our architectures. Often, when multiple phases are required, they can be generated from a single phase oscillator at a frequency several multiples higher. We found we could not create single phase oscillators with a high enough frequency. This is unfortunate as ring oscillators tend to have very poor phase noise characteristics[19]. Because of this, we typically use a large bandwidth PLL to lock on to the external references and reduce the phase noise in the clocks.

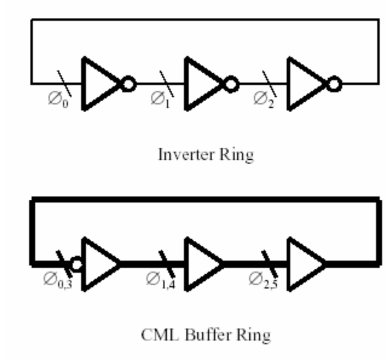

Figure 5.5: CML Ring Oscillator .vs. Inverter Ring Oscillator

The basic method of ring oscillator operation is that when an odd number of digital inverters are placed in a ring, the circuit becomes unstable with a period equal to 2N times the propagation delay in one of the inverters, (Figure 5.5). The output of each of the N inverters carries the same signal, but at a phase offset of twice the propagation delay, thus supplying an N-phase clock. According to the Barkhausen criteria, for oscillation to occur there must be unity gain around the loop and a 180° total phase shift. This gain and phase shift is provided equally by the buffers. When using differential CML circuits, inversions are accomplished by wiring, (not shown). This allows us to put an even or odd number of buffers in a ring, with a single wired inversion. The period of the CML ring is 2N times the propagation delay of one of the buffer elements, but because each of the buffer outputs can be used in an inverted or non-inverted sense, you can obtain 2N phases separated by a single propagation delay. The frequency of the ring can be modulated if desired by making the delays through the various elements adjustable using some sort of controlling signal. If the duty cycle of the phases isn't critical, a single delay element can be made adjustable, but it is more common to have all the delay elements identical.

In the original SERDES design, we used a ring VCO design conceived by Sam Steidl[20]. He came up with a method of speeding up or slowing down a ring VCO by modifying the I_{BIAS} current for the simple buffers,(see Figure 4.5), by using a variable voltage reference for the current mirror. As can be seen in Figure 4.4, the bias current affects the maximum switching speed. As the modified current mirror reference voltage reduced the bias current to slow the VCO, the VCO architecture was designated a "Current Starving VCO." This method had limitations, as the control exhibited a peaking response rather than a monotonic characteristic. In addition, the output voltage swing decreased with decreasing bias current.

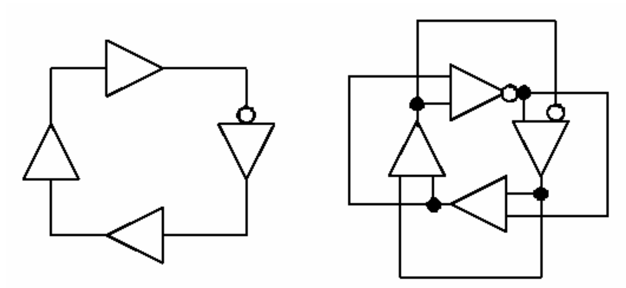

Figure 5.6: Normal ring oscillator verses a "forward leap" architecture.

In the search for better ring VCO designs, one of the discovered performance increasing methods was to use the interpolation of phases at buffer stages, shown in Figure 5.6 on the right.. Each buffer element would have two inputs, and the output signal consisted of an interpolation of the two. While one of the inputs was tied to the output of the previous element, (as in a conventional ring), the second input is connected to the output of the element preceding the preceding element. In this way, the output of an element "leapfrogs" ahead of the element immediately ahead of it. This is the type of design used in the second SERDES prototype chip. The details of the actual phase interpolation can be seen in Figure 6.10.

The VCO has proven to be one of the most difficult design challenges. Research was done into methods of designing them for high speeds, approaching fMAX. Simple bipolar ring oscillator design is discussed in [21]. A quadrature generating ring VCO at $f_{MAX}/4$ with a single phase at $f_{MAX}/2$ is described in [22], which uses a reverse-biased junction for control. The ring VCO described in [23] uses capacitors which bridge the input and output of the buffer for frequency control, and uses mixers to simultaneously increase the frequency and modulate incoming quadrature signals. A novel method of generating quadrature signals at the same frequency as a reference is given in [24]. The "leap frog" architecture mentioned in [2] achieves a higher frequency of operation over a traditional ring of buffers/inverters via the 50-50 interpolation of the phase entering each buffer with the previous phase signal. The frequency control is implemented via variable propagation delay in the buffers. The same interpolation scheme is also implemented in a fixed frequency CMOS ring oscillator in [25]. Earlier designs have also incorporated phase interpolation for speed-up in ring buffer oscillators. One of these is a VCO described in [26], where the phase interpolation ratio is variable, and is used to control frequency while allowing speed-up. However, that design was very asymmetric and 49 unsuitable for uniform phase applications such as ours. Lastly, the VCO architecture in [27] is nearly identical to that in [2], except that the fed forward phase passes through a buffer external to the ring rather than one inside it. That allows the external buffer to add delays without compromising the maximum frequency of the buffers inside the ring.

Jitter

No real world signal is completely periodic or perfectly timed. Characterization of these imperfections is called “jitter” in the time domain and "phase noise" when examined in the frequency domain. Examining individual edge placement events makes more sense in the time domain, and so "jitter" is used there. The difference in periods between two successive cycles in can be called “cycle-to-cycle” jitter. Measurement of the deviation between an ideal and a real signal period boundary is called “absolute” jitter.

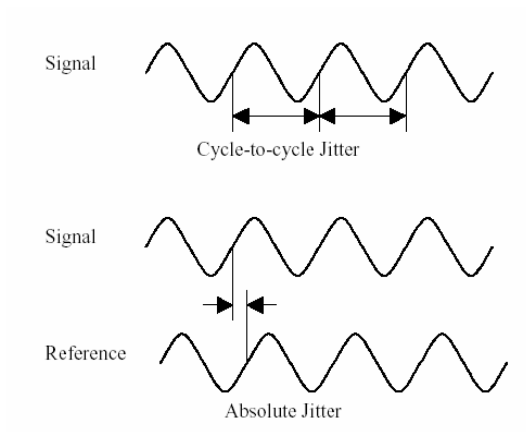

Figure 5.7: Types of Jitter

Absolute jitter is the more commonly used term. Instantaneous jitter is the measurement of error for a particular event relative to an assumed "ideal" clock. Since jitter is a measure of time, it has the units of seconds. However, for purposes of comparison it is sometimes described relative to a period of the reference signal, or a bit, in which case it can be expressed as being a unit-less fraction of such a “unit interval”, (UI).

Considering individual jitter events doesn’t easily lead to characterizations of performance, therefore aggregate values such as average, RMS, and peak-to-peak jitter are usually calculated. Peak-to-peak jitter is only meaningful for jitter that is bounded.

There are two basic types of jitter usually considered. These are random jitter(RJ) and deterministic jitter(DJ). Random jitter is usually unbounded and gaussian in nature and cannot be predicted, while deterministic jitter is bounded and can be predicted with enough knowledge of the system characteristics and the values of the bit stream. For random jitter, the RMS value is equivalent to the standard deviation, and thus uncorrelated variances can be added.

$$\sigma_{TOTAL}{}^2 = \sigma_1{}^2 + \sigma_2{}^2 + \sigma_3{}^2 + \ldots$$

Deterministic jitter can have any distribution function, including uniform distribution or discrete values. These functions are not always known. Since deterministic jitter does have well defined bounds, peak-to-peak values are specified. When calculating aggregate deterministic jitter, the peak-to-peak values can be added to calculate a worst case bounds.

Deterministic jitter can be categorized into four general types. Pulse width distortion(PWD), is usually associated with different rise/fall time characteristics of circuits and leads to longer or shorter "high" times with respect to "low" times. The use of differential logic effectively eliminates problems associated with PWD. Another name for PWD is Duty Cycle Distortion, or DCD. Sinusoidal jitter(SJ) is mostly of theoretical interest, as system performance is often tested with a sinusoidally time varying phase signal. Data dependent jitter(DDJ), is the biggest concern. This refers to the condition whereby the output of a bit is affected by the values of the prior bits. In essence, the circuit does not necessarily reach a steady state condition in a single bit time, and that state affects the next bit transmitted. DDJ is also referred to as Intersymbol Interference, (ISI). It is a direct result of bandwidth limitations in the circuit. It is possible for both low frequency and high frequency bandwidth limitations to cause this effect. For high frequency cutoffs, the signal might not reach a full high or low bit value in a single bit time, causing the next bit if different to get a "head start" towards it's final value. In a low frequency cutoff situation, a string of several bits of the same value will slowly decay towards the average value, and when a bit change does occur it has a "head start" towards the opposite value. Lastly, there are other deterministic jitter effects which are bounded, but uncorrelated with the bit stream. Jitter arising from power supply variations or crosstalk fall into this category.

Jitter can be measured directly from circuits in a variety of ways. A spectral analyzer can directly measure the standard deviation of an input signal given an appropriate trigger. For example, a clock recovery PLL output can be measured directly, or it can be measured relative to the transmitter clock. The PLL can also be tested with no input to check the spectral characteristics of the VCO. Due to the limitations of testing equipment, care must be taken when such measurements are made because often the equipment cannot sample waveforms in real time. Instead, periodic waveforms are built up over many cycles. Often the triggering event must occur dozens of cycles prior to the actual data acquisition, and this introduces an effective increase in the jitter measured proportional to the square root of the delay[28].

A single total jitter value can be calculated if a specification of allowable jitter magnitude, and the percentage of events that can violate that condition, are known. Usually this specification is based on the amount of jitter that would cause an incorrect bit value, and on an acceptable bit error rate(BER). These values can be used to find an equivalent peak-to-peak values for random jitter, which can then be added to the deterministic jitter.

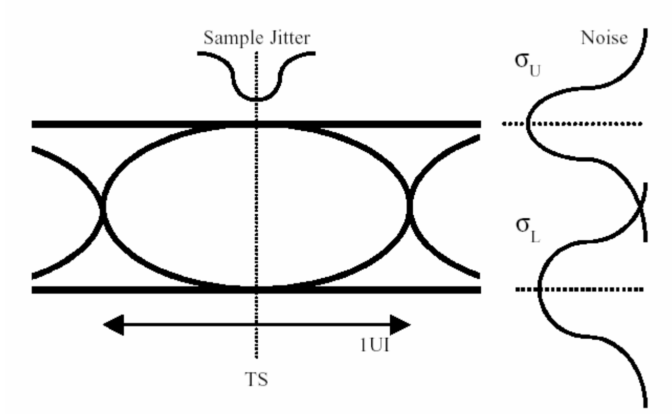

Figure 5.8: Sampling a digital waveform at a jittered time TS, with noise σ_L and σ_U.

Using knowledge of the jitter characteristics of a clock which is sampling a serial stream, the noise present in the signal, and the shape of the waveform, it is possible to directly estimate bit error rates. In Figure 5.8, a pulse eye diagram is shown along with greatly exaggerated jitter and noise histograms. The noise histograms on the left are only representative of the noise at the sample instant **TS**. If **TS** is misplaced due to jitter, the shape of the noise distributions don't generally change, but if they move closer together. There is a finite probability that a bit error will occur even in the absence of jitter, as the noise distributions overlap. The probability of a misidentified bit can be calculated as a function of the sample time. When that function is multiplied by the sample jitter distribution and integrated over time, the total probability of error can be calculated.

Phase Noise

Phase noise is the frequency domain characterization of jitter. It is more commonly used in the context of analyzing physical circuits that generate signals. The introduction of a phase noise signal into an otherwise perfect sinusoid will result in the formation of sidebands of noise near the center frequency. The result of a periodic sinusoidal phase error is a modulation, forming frequency impulses above and below the desired center frequency. If the errors are gaussian or otherwise random, they generate a noise "skirt" around the central frequency.

Most phase noise analysis starts with device noise in circuits and the resulting generation of noise at the circuit output. Individual noise sources are quantified by normalized power density with a particular distribution over frequency. Thus, the units are usually Volts^2/Hz. This power spectral density can be multiplied by the circuit power transfer function to find noise power at the output. To make the noise values more meaningful, they are usually normalized once again relative to the power of the ideal carrier output. The noise is then expressed in terms of dBc/Hz, which are decibels relative to the carrier per Hz. Specifications of phase noise have to indicate how far from the central peak that they are taken, so a typical specification might be, "The phase noise for oscillator X is - 90 dBc/Hz at a 1MHz offset from the carrier." When integrated over the circuit's noise bandwidth, the total noise power at the output can be found. Resistive thermal noise is gaussian with its variance (or spectral power density) given by:

$$\overline{V_N^{\ 2}} = 4kTRdf$$

Many treatments of noise in LCR oscillator tank circuits use only the resistive noise in calculating the phase noise spectrum. The rest of the circuit is considered noiseless. Final calculations usually end up with a spectrum that falls off as $1/f^2$. This is due to the fact that the bandwidth of the of the LCR circuit falls off as 1/f, and the power bandwidth is equal to that value squared. In practice, there is always a "noise floor" present below which the noise density doesn't fall. Also, near the carrier peak, the drop-off is usually steeper than the predicted $1/f^2$, instead being $1/f^3$.

Noise calculations for bipolar ring VCOs are similar, but they require further assumptions because there are multiple locations were noise is generated. The noise is band limited by the low pass filter formed by the resistors and the capacitance attached to the outputs either intentional or parasitic. The noise generated is independent of the number of stages in the ring[28], because the "active edge" which is traveling through the ring is present in only one buffer at a time. This is very different from CMOS ring VCOs, where the number of stages can influence the noise in a differential implementation, but not in a single ended design[19]. This is because the primary noise generating components are different, and have different distributions(white .vs. 1/f). According to [29], the SiGe HBT has superior noise characteristics when compared to normal bipolar Si BJTs, so it is reasonable to assume the resistors are still the dominant source of noise in our ring VCOs.

Another means of analyzing noise performance is based on the impulse sensitivity function, (ISF). In this type of analysis, the oscillator is considered to be a linear time variant system as opposed to a time-invariant one. The rationale for this is that the noise generated by components at different times in the oscillator period causes different effects. For example, in a circuit that has amplitude limiting, (as all oscillators must), the addition of noise when the circuit is at a peak or minimum has little effect on the output phase of the signal. The circuit tends to restore the correct amplitude before starting the transition to the other extreme. On the other hand, a noise impulse present during a rising or falling edge directly advances or retards the phase. By determining a circuit's impulse sensitivity function, phase noise can be predicted. A full treatment of this subject is presented in [19].

Clock Recovery

Clock recovery is the process of extracting a clock from a serial bit stream. For NRZ data, the process usually involves detecting edges in the serial stream relative to those in a local oscillator and adjusting the local oscillator phase accordingly. The characteristics of the loop filter used in a clock recovery PLL are dependent on the quality of clock used in the transmission of the data. A small loop bandwidth is desirable if the remote oscillator is of high quality. Any variations in edge location will be random and generated enroute. If the transmit clock is particularly noisy, a larger loop bandwidth and quick response is needed to stay aligned. Using the correct bandwidth is important as a large filter bandwidth with a good transmit clock will cause the local oscillator to track the random variations introduced by the channel, increasing the likelihood of losing lock. Likewise, a small filter bandwidth coupled with a poor transmit clock will result in a system which cannot vary quickly enough to lock to the data. The loop bandwidth also has implications for the local VCO slew rate, affecting the speed at which the system can acquire lock.

Data Retiming

Data retiming is the process of taking digital data and allowing transitions to occur at only specific points in time, usually coinciding with a an edge from a local oscillator. For complete retiming of data, a clock that has a frequency of twice the bit rate is used A master-slave latch can then ensure that edges are present only during the correct times. The requirement of a 2X clock is one of the most significant bottlenecks in monolithic transmitter design and is what prompted us to investigate lower frequency clock schemes. Normally, data to be transmitted is multiplexed to a serial stream, and then that reclocked, ensuring accurate placement of edges. Because the output passes through the same circuit, variations in process parameters don't affect the inter-bit characteristics. However, the bandwidth of the final circuitry must be large enough to prevent datadependent jitter effects from arising.

Chapter 6 Previous Work

The work, funded by the Naval Research Lab (NRL), had the desired goal of achieving a short-haul system with a 20Gbps NRZ data rate, with the possibility of perhaps reaching even a 40Gbps data rate using a process with 50GHz f_T HBTs. Current commercial designs using this same technology are at 10Gbps rates, placing the circuits well outside the realm of existing designs in terms of performance. We feel that using innovative circuits, more performance can be squeezed out of a given technology than was previously believed possible. the last fabrication run produced working chips operating at speeds in the 20Gbps range in a f_T=47GHz process. Commercial SERDES designs operating at nearly half-f_T are either extremely rare or non-existent. In order to reach these data rates under the mostly-monolithic constraint, we have had to develop new circuits, and we expect to aim even higher. To date, two complete generations of SERDES designs have been designed fabricated, and tested.

Initial SERDES Design

Researchers at Rensselaer began working on the SERDES project in 1998. Current papers in the field were used as a starting point for the research, with 10Gbps being among the fastest designs reported. The first SERDES designs from the group were submitted to be fabricated in Feb, 1999 and wafers were obtained for testing in August of that same year. The fabrication was funded by DARPA. Two separate main chips were laid out, a transmitter and receiver. In addition, several small VCO test structures were fabricated. These chips were the initial design experience and drew heavily upon[2]. They included a 10-20Gbps transmitter/receiver pair that utilized a multiphase clocking scheme to eliminate the need for a full frequency clock. The use of a multiphase clock, particularly in transmission, is deprecated due to the difficulty in creating a bit stream without an inherent duty cycle distortion due to asymmetry in the design and layout. However, given the limitations of monolithic design, we must make use of, and optimize, ring VCOs, because 40GHz single phase VCOs are not realizable in a 47GHz process.

Data "Shuffling" Scheme

In this initial design cycle, a transmitter was fabricated which incorporated multiphase VCOs in order to make use of an innovative "shuffling" data scheme that required only a quarter-frequency clock rate for operation. This extraordinary method allows us to refrain from dealing with the maximum-rate data signals until the final multiplexer/pad driver circuit, requiring only small gains in those circuits. Four 4-bit shift registers are loaded with data and fed to a 4-to-1 multiplexing circuit shown below.

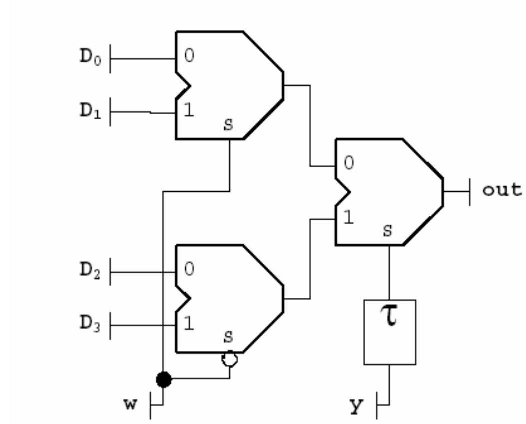

Figure 6.1: Simple 4-to-1 multiplexing scheme using quadrature phases and three 2- to-1 CML multiplexers.

The 4-to-1 multiplexer unit takes the output of the four shift registers and drives the data output pad. This was implemented using three 2:1 CML multiplexers in the hierarchical configuration shown the figure above. In that illustration, **w** and **y** represent in-phase and quadrature clocks. The input data lines on the left of the figure would remain stable while they were "selected" by the **w**-phase, That is, the corresponding shift registers would no be clocked while the data input was active. All edges present in the serial data stream are generated by the "w" and "y" phases at the latches. If the multiplexers were "ideal", the phase difference of **w** and **y** clock would determine the positions of adjacent edges in the output stream. However, these multiplexers have a propagation delay, requiring that the phase **y** be delayed by exactly the same amount. In order for the timing of the edges generated by the "y" phase to appear centered between those generated by the "w" phase, the "y" phase needed to be delayed by an amount similar to the propagation delay experienced by “w” signal from a select input, to an output of a 2:1 multiplexer. In order for this to occur, it was decided to delay **y** by sending it through the same multiplexer circuit as was used to channel the data. In the next figure, both the delay, and the timing of the data at the inputs are illustrated.

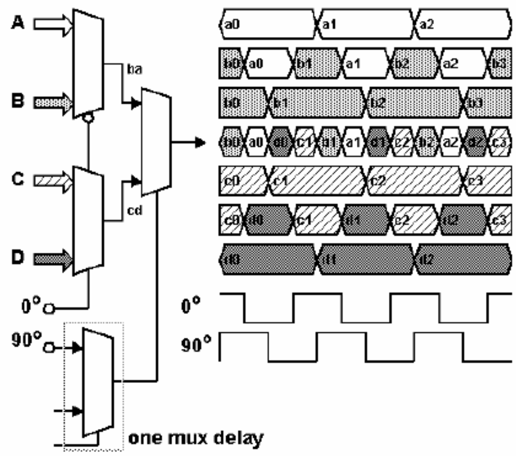

Figure 6.2: SERDES quarter-clock shuffle multiplexing. This allows correct interleaving of four data streams at the output using a quarter rate in-phase and quadrature clock.

The “shuffling” scheme is diagrammed above. The two phases, exactly 90° apart, control multiplexers to take data from shift registers A, B, C, and D. For this case, we used four 4- bit shift registers. A state machine driven by one of the VCO phases controlled when the registers would be loaded or shifted. Note how the multiplexer at the bottom left is used to delay the 90° phase signal by the exact amount of time the data takes to propagate through the latches above it. This maintains the phase difference so that the final multiplexer can introduce edges exactly between those introduced by the prior multiplexers, which handle the data. In this case, two 5GHz quadrature clocks can be used to obtain a 20Gbps output. This type of serial data generation is referred to as “unretimed”, as there is no full data rate clock latching the data after the multiplexer.

Current Starving VCO

To make the special transmit scheme operate we needed a multiphase VCO. The multiple phases were also required by the receiver architecture. VCOs are used as elements in PLLs and provide timing for all the on-chip digital circuitry. The basic ring oscillator method of operation is that when an odd number of digital inverters are placed in a ring, the circuit becomes unstable with a period equal to 2N times the propagation delay in one of the inverters, (Figure 6.3).

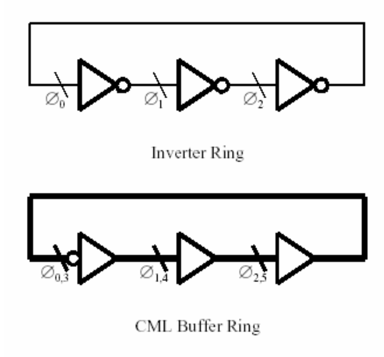

Figure 6.3: CML Ring Oscillator .vs. Inverter Ring Oscillator

The output of each of the N inverters carries the same signal, but at a phase offset of twice the propagation delay, thus creating an N-phase clock. When using differential CML circuits, inversions are accomplished by wiring, (not shown). This allows us to put an even or odd number of buffers in a ring, with a single wired inversion. The period of the CML ring is 2N times the propagation delay of one of the buffer elements, but because each of the buffer outputs can be used in an inverted or non-inverted sense, you can obtain 2N phases separated by a single propagation delay. The frequency of the ring can be modulated if desired by making the delays through the various elements adjustable using some sort of controlling signal. If the duty cycle of the phases isn't critical, a single delay element can be made adjustable, but it is more common to have all the delay elements identical.

In the original SERDES, we used a VCO design based on one conceived by Sam Steidl, a fellow researcher at Rensselaer. He came up with a method of speeding up or slowing down a ring VCO by modifying the IBIAS current for the simple buffers, (see Figure 4.5), by using a variable voltage reference for the current mirror. As can be seen in Figure 4.4, the bias current affects the maximum switching speed. As the modified current mirror reference voltage reduced the bias current to slow the VCO, the VCO architecture was designated a "Current Starving VCO". These VCOs were very fast in general, using as they did only a single buffer delay.

Transmitter

Figure 6.4 consists of a detailed breakdown of the SERDES transmitter. This design consisted of several functional units, each of which will be described. The details of the various components are given elsewhere, and only a functional overview will be presented. The pad drivers and receivers have been omitted from the diagram for clarity.

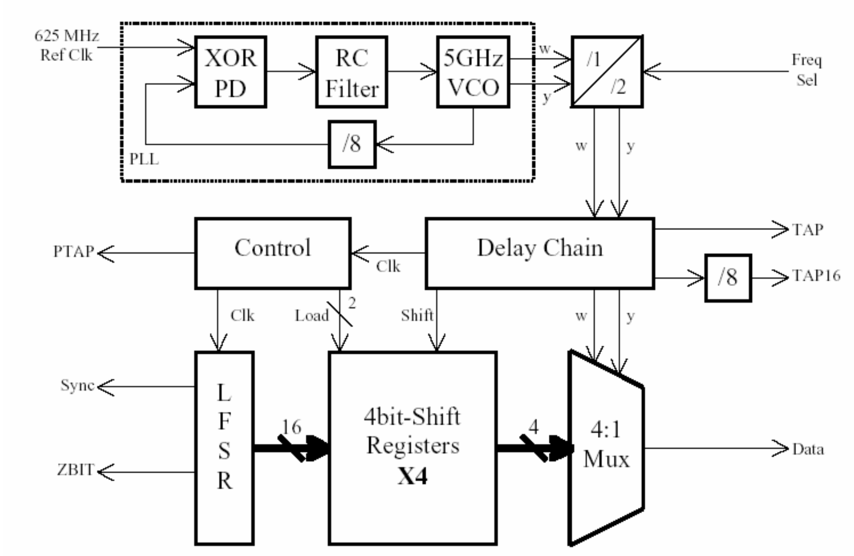

Figure 6.4: Block diagram of the SERDES transmitter.

The transmitter is designed to take 16 bits of data from the LFSR(linear feedback shift register), and serialize the data at the data output pad. The LFSR is the only source of data for this chip. Probe constraints limited the number of external inputs we could supply. The LFSR generated a repeating 15-bit sequence, which was shifted through a 16 bit register to provide data. The ZBIT line signaled when the sequence repeated for triggering purposes. A key design feature of the transmitter was that it would make use of a quarter data rate clock. When operating at a 20Gb/s transmission rate, the local voltage controlled oscillator is only running at 5GHz. The main clock for the transmitter is generated by a 4-stage ring buffer VCO, embedded in a divide-by-eight PLL, that is driven by an external 625MHz reference source. For the divide-by-8 blocks, a series of three divide-by-2 circuits were used which each consisted of a master-slave(MS) D-latch with inverted feedback. The VCO was based on a design by Samuel Steidl, another Rensselaer researcher, and used the “current starving” technique as a control method. The ring buffer nature of the VCO made multiple phases available. For this design two quadrature phases are required, and in the figure they are labeled “w” and “y”. These phases are used to multiplex the output data via the 4-to-1 multiplexer using the “shuffling” scheme previously described.

The PLL in the SERDES transmitter was the first, and it was of the most primitive sort. The phase detector was of the XOR type, implemented as a single CML gate. The RC loop filter consisted of two low-pass RC ladder stages on one side of the output of the differential CML XOR. Static phase error was not a concern in this design as the generated clock was unrelated to any other external input signals, instead, all other signals were derived from it. The external clock was expected to be a much more accurate reference, so the loop bandwidth was intended to be small, and was set to 500MHz.

The in-phase(**w**) and quadrature(**y**) signals pass through an optional divide-by-two circuit which was controlled by an external pin. This would allow either 10Gb/s or 20Gb/s operation. The division circuit generated a new **w** and **y** from the original **w** exclusively rather than dividing each of the original signals by two. It was thought that this would produce a better aligned pair. This was accomplished by using a pair of divide-by-2 blocks as described above driven in parallel by the **w** phase, but the input to one of the blocks was inverted.

The clock signals are distributed to various parts of the chip via the “Delay Chain” unit. This consisted of a series of buffers with “taps” at different points that would then pass the suitably delayed signal on to other units. In some cases “taps” were added merely to balance loading between the two phases, and for testing purposes two of these were made available at output pads. The first, labeled “TAP”, had the full 5GHz clock signal while the other “TAP16” had a divide-by-8 unit added to make frequency measurements easier.

One output of the delay chain was used to drive the bank of four 4-bit shift registers. That signal was inverted to half of the registers so that they would shift in an alternate pattern as required by the “shuffling” scheme. One bank of registers could then be reloaded while the other bank was in the process of shifting.

In order to generate the clock for the LFSR and the “load” signals needed by the shift registers, a state machine (CONTROL) unit was necessary. This circuit generated a 25% duty cycle pulse at one fourth the frequency of the main phases, which was then used to enable a load for one half the shift registers. The other load line was derived from the first, but changed after a delay of one half of a main phase period. The LFSR also advanced to the next state using this signal. An additional signal equivalent to a divideby- 4 of the main clock was incidentally generated, and sent to the PTAP output pad to monitor the operation of the control unit. The output of the shift registers was sent to the final 4:1 multiplexer to generate the final serial stream. In simulations without parasitic extractions, the transmitter performed up to 23Gb/s.

Receiver

The design of the SERDES receiver is diagrammed in Figure 6.5. The receiver IC consists of two major sections, concerned with test signal generation and the receiver itself. The upper part of the figure is for test signal generation and features a 5GHz VCO of the simple current-starving architecture. This VCO is controlled by a voltage supplied via an external pad. Three of the VCO phases pass to a frequency modifier block used to create the onboard generated test signals.. One phase of the VCO is available at an output pad for monitoring. In the frequency modifier block, one phase is passed through a divide-by-two circuit identical to those used by the transmitter. Two of the phases that are in quadrature are combined using a symmetric XOR to produce a signal at twice the VCO frequency as described in[22]. One of the three phases is passed through unmodified. These signals, at the VCO frequency, half the VCO frequency, and twice the VCO frequency, are fed to the input select unit. The input select unit also receives one signal from an external pad. The selector is just a 4:1 multiplexer controlled by two external select signal pads, (SelA,SelB).

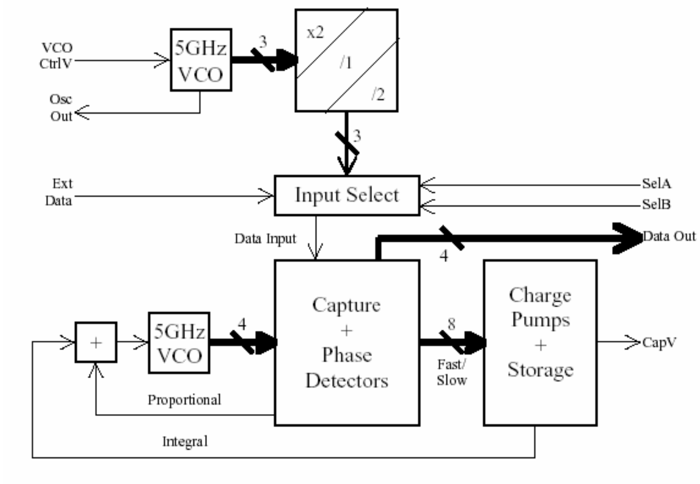

Figure 6.5: Block diagram of the SERDES receiver.

The output from the input select unit demarcates the boundary between the testing circuitry and the receiver proper. The receiver has a separate VCO that is used to drive the data and phase detector latches. Each of the four VCO phases and their compliments are used, resulting in eight samples taken per VCO period. Four of these samples are intended to occur in the center of data bits while the others are used to detect the position of the potential edges occurring between the bits. The values captured by the data latches are sent to four “Data Out” pads as is shown in the figure on the right. The exclusive-OR between each data bit and an adjacent edge detection sample would indicate the presence of an edge. Edges detected just after a data bit rather than just before would be an indication that the edge was early, and a “slow” signal would be generated as a command to the VCO. Likewise, if the edge arrives just before a data bit, a “faster” signal is generated. In this design, the phase detector XOR outputs were gated so that they were only active immediately after their second input became valid. The eight potential fast/slow signals are combined to produce a proportional signal that is sent to the VCO. In addition, they are used to control charge pumps, implemented with FETs, connected to a capacitor to integrate the error. The proportional and integral signals are then summed and used to control the VCO, closing the feedback loop.

Layout

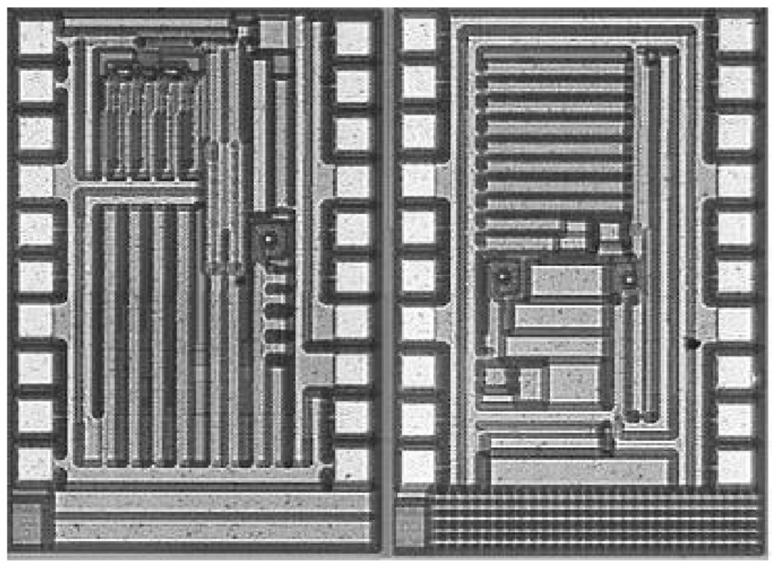

Figure 6.6: Microphotograph of SERDES I chips. The transmitter is on the left and the receiver is on the right.

The transmitter and receiver were manufactured as shown in Figure 6.6. Each of them was approximately $2mm^2$. Both chips were meant to be tested via a pair of 10 pin, (6 data channel), probes. In addition, a slightly modified layout was created which was intended to allow data generated by the transmitter to pass directly into the receiver while still allowing testing using the same pair of 10 pin probes.

Design for Testing

The limited external signal probe capability prompted us to incorporate an LFSR to generate pseudo-random data for the test transmitter to send. Due to the limited testing facilities, we used on-chip signal generation to provide the chip with data, and the data was only demultiplexed to 4 lines. The final 4-to-16 stage would be implemented in a later design. Both chips deviated significantly from expectations based on simulation.

The results from this initial design cycle showed how far the simulations deviated from the physical circuits, especially in the area of the performance of the VCOs. Analysis led to feedback with IBM's modeling staff to improve accuracies. In testing, the receiver operated at nearly the desired rate, while the transmitter under-performed by around 25%. A paper was prepared and submitted to the International Solid State Circuits Conference (ISSCC), but unfortunately was not accepted for publication, most likely due to the mismatch in operating ranges of the transmitter and receiver.

The transmitter VCO center frequency was intended to be 5GHz, with a control voltage between -.8V to –1.6V, so at –1.2V we should have approximately 5GHz. The gain even in simulation was non-linear, and actually went negative in testing at high frequencies due to drive beyond peak fT current. We decided to come up with a more linear and controllable design for the next prototype.

The transmit multiplexer scheme was found to be less than ideal as well, and showed a noticeable duty cycle. This was partly due to asymmetries in layout, and also due to the multiplexer scheme itself. We had intended to delay the second phase by the same amount as the first, but there were additional unplanned for propagation times we had not considered. One of the main problems was that the CML 2-to-1 multiplexers had a different propagation time to output for the data inputs as compared to the select line. We were unable to achieve a perfect balance using these multiplexers. There was always a constant offset between edges produced by the particular phases. This was the impetus for much of the redesign in the next prototype.

Second Prototype

Using the information learned in the first design cycle, we began to develop a second version of the serializer/deserializer chips in late 1999, early 2000. The complete new SERDES II design was created and submitted for fabrication in March of 2000. The window of opportunity for fabrication was made suddenly available by a corporate benefactor, Sierra Monolithics Inc. Due to the hurried nature of the preparation for tape-out, and because the allocated chip size changed more than once, we were unable to do the kinds of extensive verification of the final layouts that we normally would have done. Because of this, several small errors were able to make it into the final layout, resulting in small degradations in performance of the physical chip.

The new design used a 3.4V, (down from 4.5V), nominal power supply, and was created using a 5 layer metal process. Most of the advantages of working with 5 layers were lost due to the funding organization’s desire to have this design be packageable. Towards that end the chip was designed with both regular bondpads which we could probe using the test station, and C4 solder bump pads for flip-chip packaging. A single chip was created for both transmit and receive operations.

In July 2000, the chips were received and testing began. During the entire period between design revisions, IBM was engaged in refining their process, and the models along with it. This has lead to more small deviations from expected behavior, but these deviations are smaller with this revision than they had been in the previous one.

Symmetric Multiplexer

To address the problems with the first prototype transmitter duty cycle, we designed a new 2-to-1 multiplexer with characteristics designed to give us more uniformity in propagation delay for the data and select inputs. This new multiplexer circuit which was intended to alleviate the imposed duty cycle on the output by the previous design. Prior to this, the multiplexer had used a multi-level approach using a regular CML tree. This lead to asymmetry in output timing due to the different propagation times of the different signal levels through the CML circuit. The circuit below is much more symmetrical.

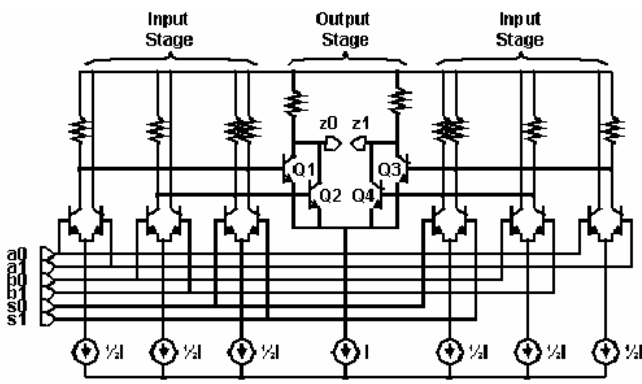

Figure 6.7: Symmux circuit designed to have equal propagation time from each of the inputs to the outputs.

This new multiplexer circuit features a *nearly* uniform propagation delay between any of its three inputs. It is used in both the final multiplexer, as well as in the dummy delay multiplexer intended to delay clock phases by the exact same amount of time, to ensure symmetric output. A provisional patent was obtained for this circuit, (Figure 6.7.) It should be noted that it was not 100% data-select symmetrical due to minor loading effects in the "interior" of the circuit.

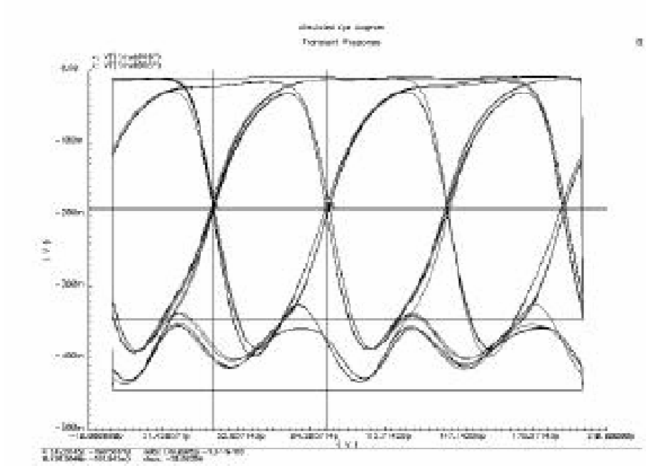

Figure 6.8: Duty cycle plot of symmux based 4-to-1 multiplexer under ideal conditions.

Using this circuit and extremely careful and symmetrical layouts, we are able to get very reasonable performance. The simulated plot above, Figure 6.8, shows how much duty cycle the multiplexer scheme imposes when fed with correctly aligned signals. Measurements of the above plot show a very small duty cycle.

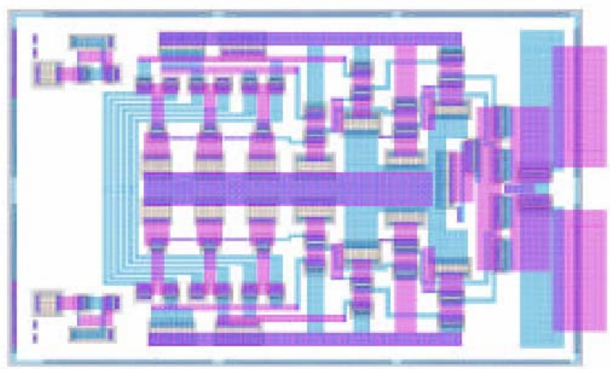

Figure 6.9: Symmux layout. The symmetry of the layout ensures that parasitic effects are matched as closely as possible.

Figure 6.9 is a layout of the final 2:1 multiplexer in the transmitter, which directly drives the output pads. Similar attention to detail was used throughout the critical areas.

Leap&FFI VCO

For performance reasons we decided to come up with a new VCO architecture for the second prototype. The current starving VCO was too non-linear to safely use in the PLLs. More complex designs seemed to always degrade final performance to unacceptable levels. Eventually, a match of circuit speed-up techniques and control-enhancing slow-downs was achieved which yielded acceptable performance.

In the search for better ring VCO designs, one of the discovered performance increasing techniques was to use the interpolation of the phases at buffer stages to speed up the effective propagation delay at each stage, (See Figure 5.6.) Each buffer element would have two inputs, and the output signal consisted of an interpolation of the two. While one of the inputs was tied to the output of the previous element, (as in a conventional ring), the second input is connected to the output of the element preceding the preceding element. In this way, the output of an element "leap-frogs" ahead of the element immediately ahead of it. We chose a four buffer VCO architecture as being the simplest leap-based design that could provide us with the quadrature phases. Figure 6.10 below shows the basic leap architecture operation. This design utilizes the feed-forward and averaging scheme to allow each buffer to "anticipate" the incoming edge, resulting in a 33% increase in operating frequency. The phases p_i would normally be separated by a full stage delay. However, because p_n is based on the average of the two prior edges, the output can be anticipated. As described above, the frequency of the VCO was controlled by varying the amount of current through the CML buffers, altering the current-dependent value of f_T.

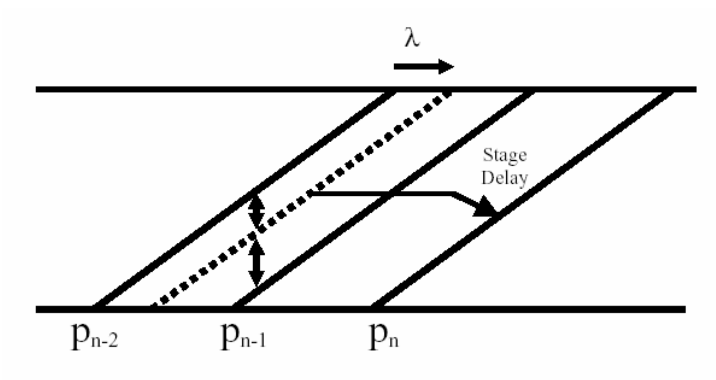

Figure 6.10: "Leapfrog" architecture interpolation plot. By averaging the signals from previous buffers, a 33% increase in frequency can be obtained.

This technique gave a great performance boost but didn't address the controllability issue. What was needed was a way to speed up or delay each phase without sacrificing higher frequency operation. Normal techniques such as using varactors were considered, but abandoned as they either degraded performance to too great a degree or had too small a range.

Eventually, the idea of the FFI(Feed Forward Interpolated) VCO was developed, in which the amount of interpolation between the stages can be varied, (λ in Figure 6.10), effectively "sliding" the position of the interpolated edge. The circuit diagram for one of the buffer stages of the new VCO is shown below in Figure 6.11. It was used in both the transmitter and receiver circuits.

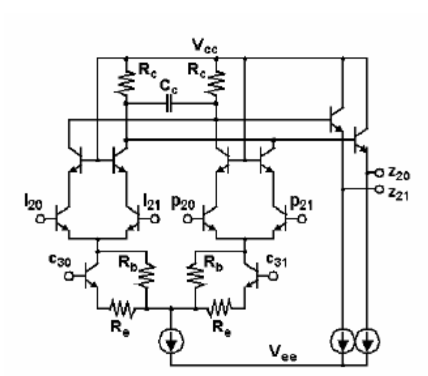

Figure 6.11: Buffer element of the Feed Forward Interpolated(FFI) VCO. By varying the interpolation ratios, the frequency can be controlled.

This VCO phase interpolation scheme would exhibit a great deal of jitter as a result of attempting to interpolate between the clock phases if the edges were steep. Because of this, the buffers were intentionally linearized and the gain reduced to an extent to allow sinusoidal waveforms, and resulting in fairly smooth interpolation. The purpose of the R_e resistors in the CML tree paths in the circuit above are to provide this linearization. In addition, resistors R_b prevent a 100% current swing to one side or the other. The control voltage has to be supplied differentially, and the circuit exhibits a high degree of common mode noise rejection. The optional capacitor at the top of the circuit allows control of the center frequency. In the circuit's highest intended frequency incarnation, the capacitor is left off completely.

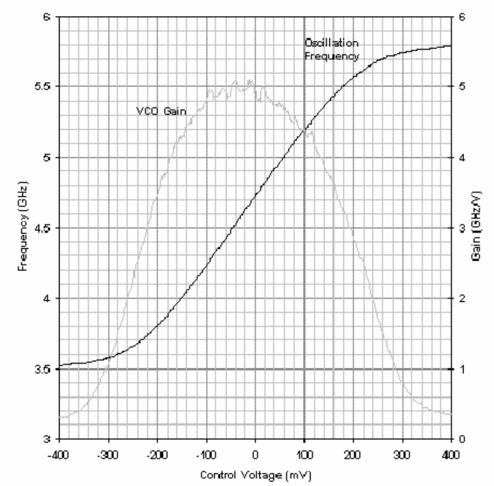

Figure 6.12: FFI VCO control plot. The plot above shows a wide linear range.

The plot above shows the large usable linear region of the VCO, as well as it is large range. The VCO was also linearized to provide a more sinusoidal output, as slowly rising edges can be interpolated with a greater degree of control. The phase noise was very good compared to other ring VCO circuits, being –90dBc/Hz at 1MHz.

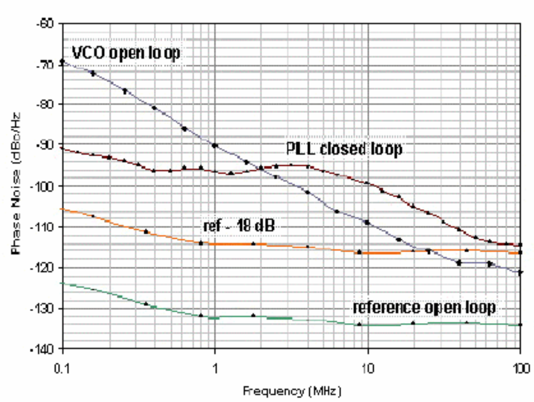

Figure 6.13: Closed loop performance of the FFI VCO in the PLL.

The phase noise of the FFI VCO is shown in Figure 6.13 for closed and open loop operation. The PLL was tuned to minimize output noise based on simulations of the VCO, and the noise characteristics of the 625MHz reference source[30]. Because of the characteristics of the source were less than ideal, the loop bandwidth was reduced to compensate. Below is a VCO layout showing the degree of symmetry achieved, (Figure 6.14).

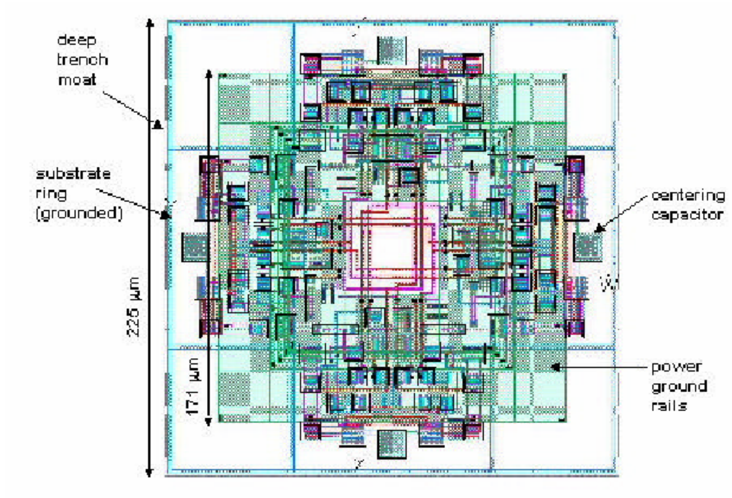

Figure 6.14: Layout of the FFI VCO showing symmetrical loading and isolation.

With a better reference signal source available, we can increase the loop bandwidth, making the transmitter track the better external source further out.

Transmitter

The major performance changes to the transmitter architecture over the original SERDES design are the introduction of the symmetric multiplexer, and the use of newer FFI VCO design. The new transmitter design, (Figure 6.15), was given multiple VCOs and a variable divider inside it's PLL to allow it to transmit at several different frequencies. By varying VCOs and dividers, 10Gbps, 20Gbps, and even 40Gbps could be selected as desired output rates, given a nominal 625MHz input signal. The 625MHz signal could be varied through a large range with the transmitter still remaining in lock, allowing almost any intermediate data rate. Circuitry for testing now included a 12-bit pseudo-random generator in the transmitter to generate 16 streams of pseudo-random data.

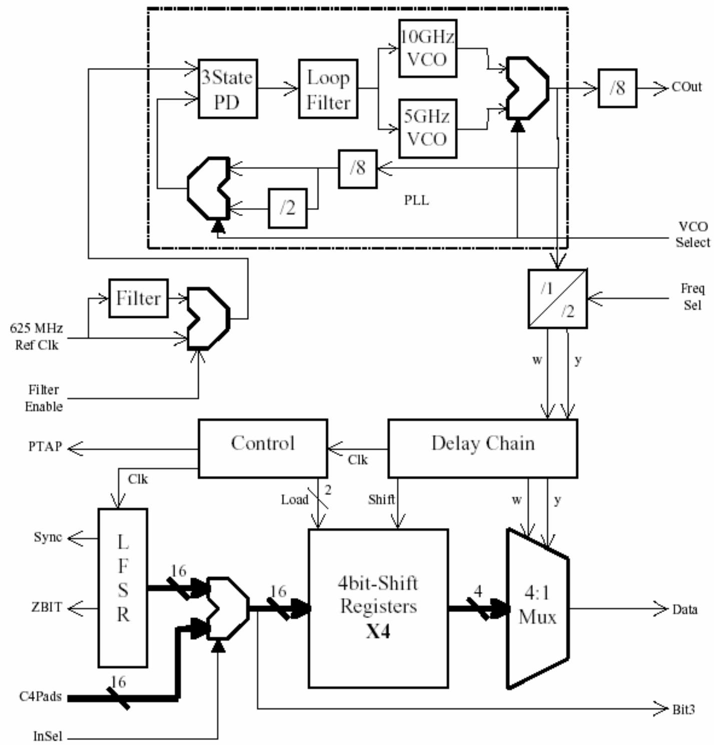

Figure 6.15: SERDES II transmitter block diagram.

The transmitter PLL also featured a new active op-amp based loop filter, and moved from an XOR phase detector to a 3-state phase detector to more closely track the **RefClk** input. Much more attention was given to optimizing this filter to ensure the best possible performance with regards to lock in range, lock in rate, and minimizing output phase noise. In the original prototype, only four bits were used for testing. In contrast, this design features a full 16-bit architecture. A multiplexer was added to choose between the LFSR data and external data that would have been delivered via the C4 pads. Although this revision didn't quite achieve 40Gbps, it did generate outputs well over 20Gbps with the limiting factor believed to be in the timing and pseudo-random generation circuitry, rather than in the main transmitter itself.

Receiver

The first part of the receiver, (Figure 6.16), is largely identical to the design in the first prototype. The input select lines can be used to choose among three internal VCO generated periodic signals, or the external data input. Like the transmitter, the receiver benefited from a new active loop filter, and used an op-amp based filter/integrator to replace the more problematic charge-pump design. Also, a simplified phase detection scheme was incorporated that reduced the delay between the sampling used for detecting the phase, and the signal passed on to the filter. A final 4-to-16 demultiplexer was added to enable output to the full 16 parallel lines. Also introduced was a 12-bit state machine recognizer to match with data from the transmitter's LFSR, intended for the detection of true bit error rates (BER).

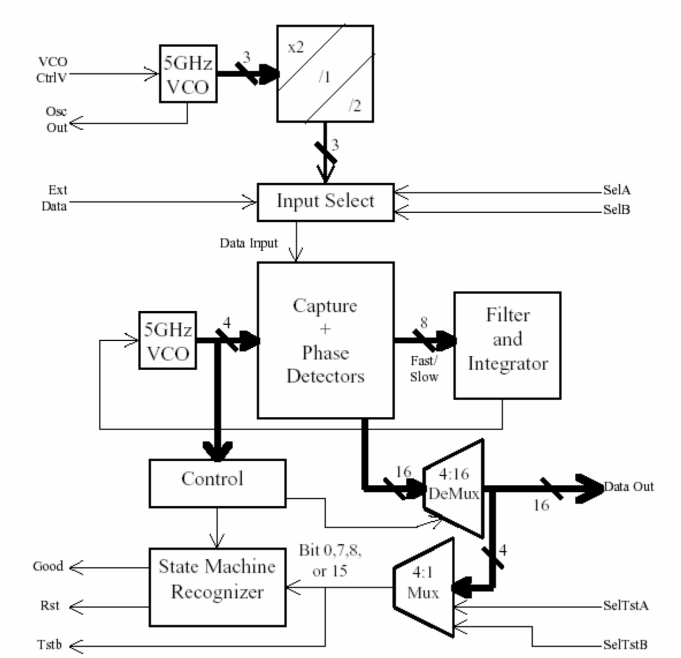

Figure 6.16: SERDES II receiver block diagram.

This BER testing circuitry allowed one of four parallel output bits to be selected, (via **SelTstA** & **SelTstB**), and sent to the LFSR-based state machine recognizer. This selected bit could also be monitored on an output pad, **Tstb**. On a bit miss-match, the recognizer would reset the LFSR and this would be detected on the **Rst** line. If the recognizer passed through an entire sequence of 4095 bits, the overflow would cause the **Good** output to be toggled. At 20Gbps data rates, this would generate a square ~150KHz output signal, easily detectable/countable by the equipment.

Layout

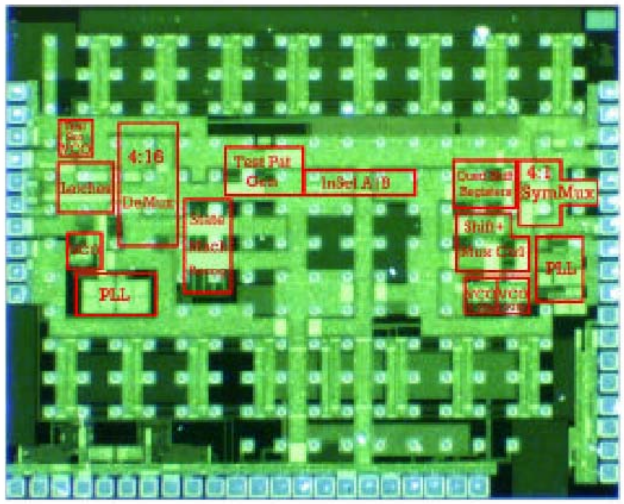

Figure 6.17: Layout of the SERDES II prototype 20Gb/s chip.

The SERDES II chip seen above in Figure 6.17 has C4 pads as well as wirebond, as it was envisioned that it might eventually be packaged. We added the necessary circuitry to handle a full set of 16 parallel inputs and outputs via the C4 pads, leaving only a few of the signals accessible over the bondpads, which we could probe prior to considering the chip for packaging.

The total chip size was approximately 12.5mm^2 in size, however much of this large area was unused by the serializer, deserializer, and test circuits. The large size was required for the necessary C4 pads, and some of the unused area was appropriated for the test circuits of other group members.

Results

Several errors were made in the design of this chip which might have been caught had we had more time for testing and layout. A small mistake in the receiver loop filter greatly reduced its lock in range. (A node in between two capacitors was grounded when it should have remained floating.) Parasitic effects once again had an impact on the VCO center frequency that did not show up in simulation. With these problems we still achieved transmit capability of 14.27Gbps to 21.58Gbps, and the receiver worked from 16.8Gbps to 18.5Gbps.

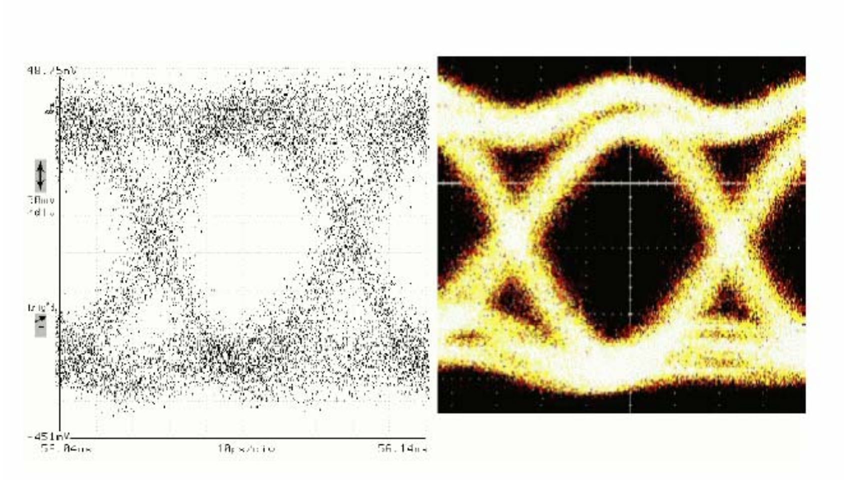

Figure 6.18: SERDES II 20Gb/s eye diagram on the left, compared with a commercial 20Gb/s eye in a technology with twice the effective performance.

We were able to rent test equipment and obtain the spectral VCO data presented earlier. In addition, we used the sampling scope to produce numerous eye diagrams. In Figure 6.18, the eye on the left was directly captured output from the transmitter at 20Gbps. The right eye was produced by a commercial group using 7HP technology that has a much, (over 2x), higher f_T. We obtained a 400mV output swing on each line of a differential pair yielding a difference swing of 800mV.

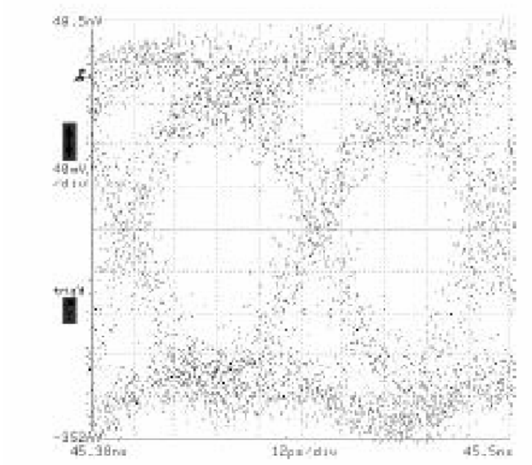

Figure 6.19: Dual eye diagram showing small duty cycle.

When a dual-eye plot of the transmitter output is taken, duty cycle is only *minimally* apparent, as is seen in Figure 6.19. Thus, the new symmetric multiplexer has gone a long way towards alleviating problems with the "shuffling" quarter-frequency transmission scheme. However, we later found that due to small design errors, the performance could have been even better.

The results regarding the VCO, transmitter multiplexer scheme and circuits were written up and submitted to the IEEE Journal of Solid-State Circuits (JSSC) for publication. As of this date it was not accepted, possibly again due to the mismatch of transmitter and receiver rates. We are seeking the opportunity to publish the bulk of this research elsewhere, after careful review and re-edit.

Second Prototype Corrections

Various problems were identified within the second prototype, almost all of which are correctable in future revisions. Some of these would have been dealt with before fabrication had there been sufficient time for more testing. We were again somewhat disappointed that the VCOs under-performed the simulations, and have adjusted the simulation techniques accordingly.

In the receiver, as mentioned earlier, a loop filter error reduced the lock in range. Also, there were several areas where loading effects caused improper operation at higher data rates. When the designs were re-simulated using more accurate parasitic modeling, interconnect capacitance apparently was such that some line drivers were unable to drive the lines with sufficient speed to ensure correct operation at or near the upper limits of the design specification. This was especially apparent in the 4-to-16 receiver demultiplexer, where each driver had to drive four others as well as all the interconnect, (which turned out to be quite a lot.) The circuitry to handle 16 bits takes up quite a lot of real estate, at least when compared to the previous designs. It may have been mistakenly assumed that there would be intermediate buffers, as these blocks were laid out by different researchers. A similar loading problem also occurred in a delay chain that drove the LFSR in the transmitter. Above a certain frequency, the LFSR produced no output. Simulated at low speeds, or when modeled without parasitics, the circuits performed correctly. In order to fix these problems, the necessary undersized drivers were modified, and the circuits tested using full parasitic simulations at the desired working speed.

Other problems that have been identified include a miss-match in the circuits driving the final transmitter multiplexers. For complete uniformity of operation, all inputs should be driven at the same common-mode level and by emitter follower transistors of the same size. It was found that the outputs of the four 4-bit shift registers, (which feed the final 4- to-1 multiplexer circuit), used 1μm transistors while the clocking signals to the select lines were 4μm. The clocking signals were split giving an effective driving capability of 2μm for each, so the outputs of the shift registers should have been 2μm as well. In addition, the transmitter also had a hard-wired constant value connected to a “dummy” multiplexer that was intended to only provide a delay. This constant was also at an improper level. As with the other problems, work-arounds have been developed and simulated.

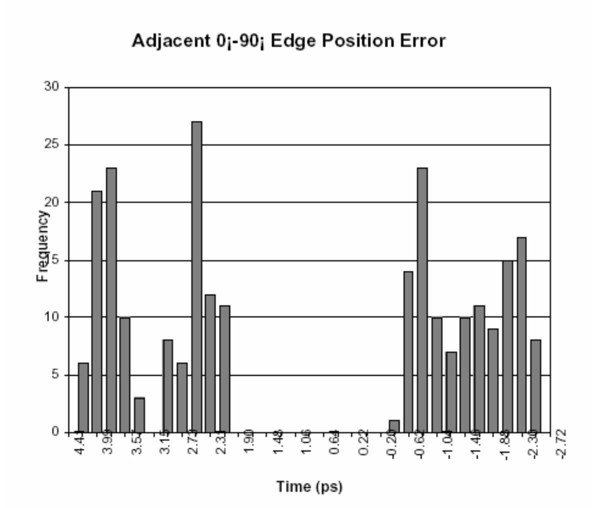

Figure 6.20: Simulated edge position frequency distribution for corrected SERDES II under ideal conditions.

All together, these errors in the symmux led to a visible duty cycle. The plot above shows predicted behavior of a "corrected" symmux under ideal conditions. The average error is less than 1ps, which is far superior to that of the original prototype employing regular CML multiplexers. After "correction", the standard deviation of the error has been reduced by 45% to 2.43ps, when compared to the "uncorrected" circuit that was actually fabricated. However, this standard deviation is still much larger than that of the simple CML multiplexers and so is far from ideal. The spread is mostly due to the reduced bandwidth of the symmux giving rise to data-dependent jitter. Other problems with the circuit are it's potentially irregular rise and fall times. Because pairs of independent transistors act on each output line, conditions can arise when they are not in the same state, or not even in a fully differential state when compared to the transistors on the complementary output line. Since some of the preliminary buffers have different internal loading, the propagation time for the select line is still slightly different from that of the data lines.

Some of the most puzzling deviations from ideal behavior were investigated to see if they were due to data-dependent jitter. It was initially found that the "corrected" circuit still exhibited multiple peaks in the distribution for the odd edges. To better visualize what was occurring, we hit upon the idea of using a phase plot, distributed over four bit times, so that the edges are automatically partitioned by the hardware that generated them, (Figure 6.21). You can see the output below, which is quite interesting. This is a 76 standard XY-cartesian-polar figure, with zero degrees on the far right. Ideally, all edges would occur on an axis. Bit A occurs in the upper right quadrant. Bit C is in the lower left, etc. The A-B and C-D odd edges with their two error peaks are clearly visible at the top and bottom.

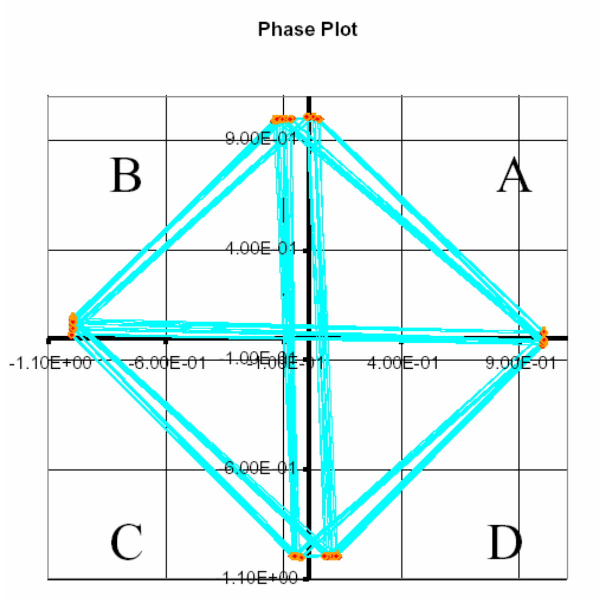

Figure 6.21: Plot showing edge placement errors.

It is quite interesting that there is a *complete* absence of almost vertical edges in an X-pattern connecting opposite peaks on top and bottom! This indicates that if an odd edge is early, the odd edge after it is **always** late. The converse is also true. If an odd edge is late, the odd edge after it is always early. In order to determine if the edge delays were dependent on the bit values immediately preceding them, the same data was plotted, but the radial distance was varied based on the number of identical bit values preceding the edge. This is shown in Figure 6.22 below.

Phase Plot

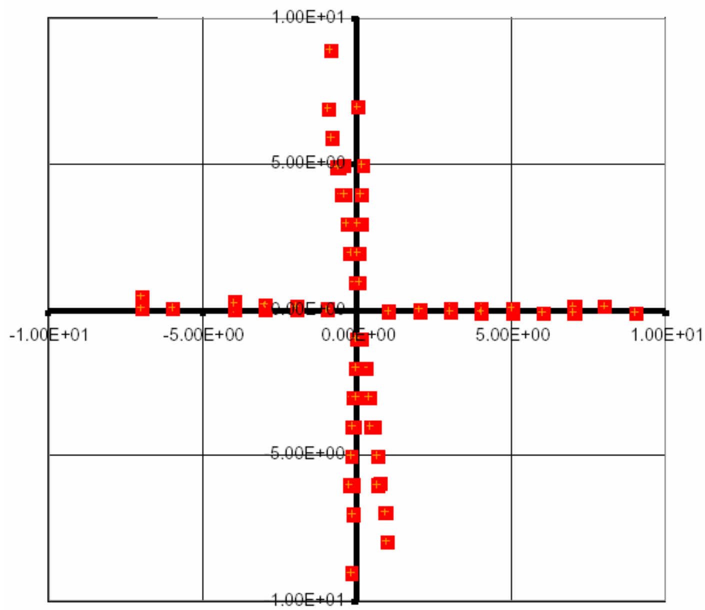

Figure 6.22: Radial plot showing dependencies between data and edge placement error.

In this plot, the edge placement error is the angle between the point and the nearest axis. If the error grew with longer sequences, the edge marks would tend to curve away from the axis as the radial distance increased. Alternately, if the errors were due to rapid alternation of bit values, the edges closest to the origin would be furthest from the axis, and the edges would curve back towards the axis as radial distance increased. As can be seen, the edges appear along radial lines indicating there are virtually no dependencies on the length of same-bit sequence preceding the edge! This was quite unexpected.

The edge data was finally analyzed and explained using a C-program written that correlated each edge error with the hardware that generated it, the sign of the error, and the full n-bit pattern that preceded it. It was found that there was a strong disparity between a bit sequence and an identical except complementary sequence. For a fully differential circuit this should not be possible.

The difference in 0/1 behavior was finally tracked down to self-heating effects[14]. In the multiplexer architecture, we used "dummy" multiplexers to match loading and delay, and these used fixed value "constants" at some of the inputs. The adverse effect of the constant was to bias one side of a differential pair more heavily than the other, increasing it's local temperature and altering it's behavior with respect to rise and fall times. The self heating time constant modeled the SiGe technology is on the order of microseconds, which corresponds to tens-of-thousands of bit times. This has important implications for simulation, as the initial temperature is derived from the initial conditions for a transient simulation. To get accurate measurements, the differential test input voltages should be zero at t=0 to prevent an initial temperature bias that will last for thousands of bit times.

One last area of potential problems in the second prototype to be addressed was in the transmitter state machine, (**Control**), which initiated the loading verses shifting operations of the four shift registers. This state machine used a single gate delay feedback path with master-slave latches for all the state bits except one, which needed to be a half-period delayed copy of one of the other outputs. This was accomplished by inverting the clock to a master-slave latch driven by the output to be copied, but this created a half-period propagation criteria whereas the rest of the circuit could make use of the full period. By re-implementing the last bit using combinatorial logic from just the previous state, and retaining the inverted clock, the potential top frequency of this state machine was effectively doubled.

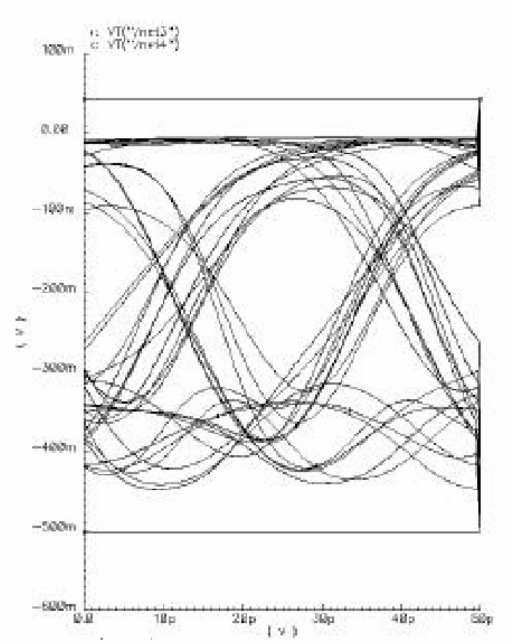

Figure 6.23: Simulated SERDES II multiplexer output at 40Gb/s under ideal conditions.

Figure 6.23 is a rough simulation of how the second prototype SERDES output multiplexer might look at 40Gbps. The output was not physically achieved, but there are no fundamental reasons why we should not be able to reach this goal in future designs. Of course more work does need to be done to “tune” the designs for operation at these frequency extremes. The complex circuitry of the symmux decreases the bandwidth of the output circuitry to the point where it introduces a large data-dependent jitter, which would seem to make reliable operation impossible using that design. Future work will have to incorporate finding a way around this limitation, as well as reducing or completely eliminating the duty-cycle from the transmitter output.

Chapter 7 Current and Future Work

After the second prototype had been evaluated, and “corrections” to the design proposed, work was done on improving some of the 20Gbps features such as the lock-in time of the receiver. Simulations were carried out to test various new ideas in filters, phase detectors, and whole PLLs.

Pad drivers and receivers have been redesigned to allow for easier testing. (For example, some of the low frequency signals used to require a DC offset which is undesirable from a testing point of view.) Termination characteristics are more accurate. The transmitter PLL bandwidth can be increased as necessary to better take advantage of the newly available high accuracy reference signal source.

The PLL loop filter in the receiver has a fairly narrow bandwidth, which is desirable from the perspective of suppressing noise in the input data. However, this prevents the local oscillator from shifting to a different frequency in a relatively short time. In a continuous data streaming system, such as SONET, this would not be as important as in a packetized network where it would harm performance. For this reason, a frequency detector should be added to cause a more rapid transition to the target frequency. The frequency detector can be incorporated into an n-state phase detector with little additional circuitry penalties.

40GBps

We have gained valuable experience in designing circuits in SiGe, and identified potential problems that reduce performance and quality of the output. The target of this new design cycle will focus on achieving 40Gbps in a 50GHz fT process, as well as preparing to move to ever faster technologies such as 7HP. As in all designs, certain tradeoffs are necessary to achieve target objectives. Just as gain and bandwidth are traded in amplifier designs, so must be data rates and output signal quality. Ultimately, the desired bit error rate of the application sets a lower bound on signal quality. It was felt that achieving the 40Gbps data rate was a new and significant hurdle to aspire to. Focus went immediately to the most obvious bottleneck, the symmux. As was shown in the figure in the last section, the bandwidth of the symmux was far too limited to allow 40Gbps data through without introducing huge data dependent distortions In addition, we hope to retain the flexibility to transmit at various other rates.

SymGate

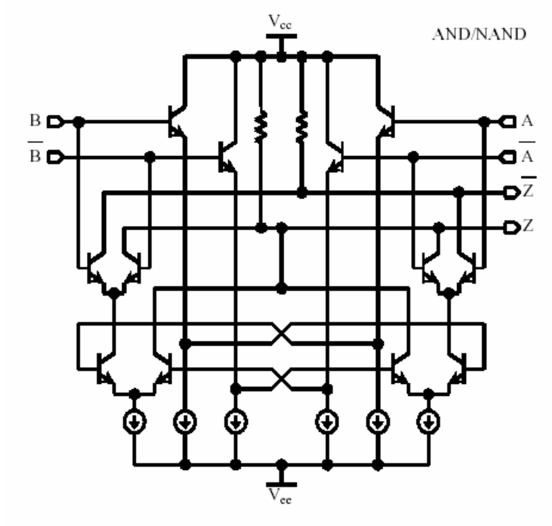

Figure 7.1: Completely symmetric two input CML gate.

Ideas were developed and new circuits tried, including several new types of symmetric gates. The circuit in Figure 7.1 is a perfectly symmetrical AND/NAND (or equivalently OR/NOR) gate. Much thought was put into ways in which gates like these could be combined to arrive at a balanced multiplexer with high bandwidth. Unfortunately, the best design we arrived at had double loading on the select lines as compared to the data lines, and was thus not ideal. These gates are still useful building blocks and may be used in other portions of the circuits where even timing is critical.

Edge Steering Multiplexer

Offsets between edges in the output serial stream continued to present the largest problem. The combination of symmetric circuit requirements to avoid duty cycle, and high circuit bandwidth to avoid data dependent jitter was difficult to overcome, however the following circuit should alleviate this problem to a great extent.

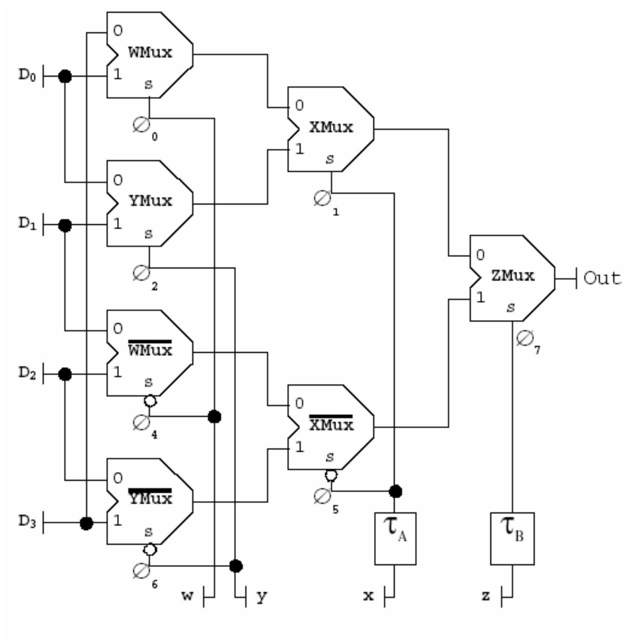

Figure 7.2: Proposed new multiplexer architecture. This edge-channeling multiplexer creates all edges that will appear in the output serial stream at the first tier. These edges then propagate with a uniform delay through the rest of the circuit resulting in an output stream with no inherent duty cycle distortion.

The multiplexer architecture shown in Figure 7.2 has several improvements over previous generations. In the first SERDES chip, the propagation delay was addressed to only the first order. High bandwidth multiplexers were used, but the output possessed a duty cycle. The SERDES II chip improved on that using the symmux architecture, but suffered from poor bandwidth and lack of complete symmetry. This new "edge channeling" multiplexer above uses the high bandwidth characteristics of standard CML multiplexers while addressing the symmetry issue. It's operation can be understood by referring to Figure 7.3 and Figure 7.4 which show timing diagrams of the clock phases and the data at the multiplexer outputs respectively.

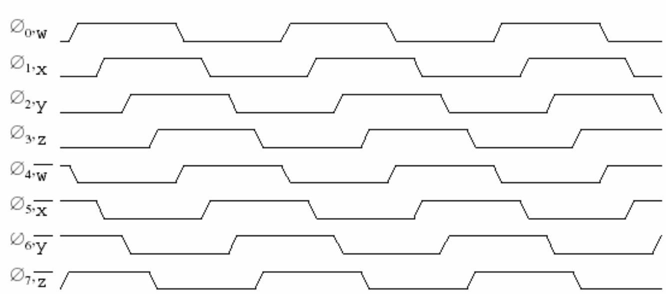

Figure 7.3: Eight phase clocks as generated by a differential four buffer ring VCO. Both phase numbers and letter names are used to refer to specific phases.

Using a differential four buffer VCO, we obtain eight effective clock phases, as diagrammed in Figure 7.3. As can be seen, each phase has a matching complementary phase and we refer to them at different times by their letter or number designations.

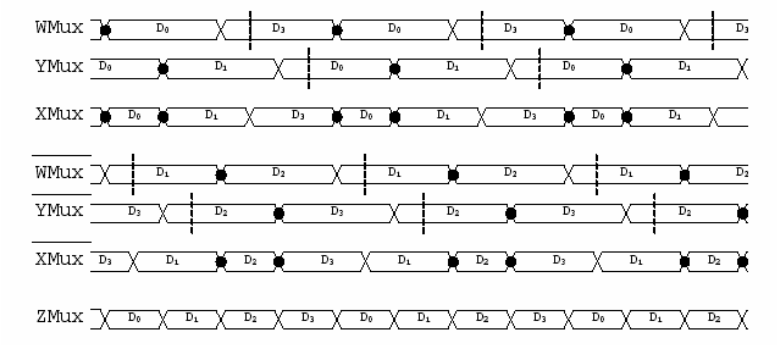

Figure 7.4: Channeled edge architecture multiplexer outputs. Each signal represents the output of a 2:1 CML multiplexer, and is labeled on the left with the phase line used to drive the multiplexer select.

The multiplexer operates by using the top(leftmost) tier of 2:1 multiplexers to generate the edges that will be present in the final output. This topmost tier is driven by the “w” and “y” phases and their complements. For example, a potential edge exists between data bits zero and one, (D0 and D1 in Figure 7.2). Therefore, both these bits are present at the input to a multiplexer, in this case the one labeled “YMux”, which is driven by the “y” phase. On the rising edge of the “y” phase, if the values of bits D0 and D1 differ, there will be a rising or falling edge at the output of the “YMux” multiplexer. Similar edges are generated for each pair of potential edges in the final output sequence. These edges are indicated in Figure 7.4 by the presence of a dot.

Once all the critical edges are generated by the top tier, the rest of the multiplexers are used to “channel” these edges to the output. In the second tier, the multiplexers are driven by the “x” phase and it’s compliment. When the rising edge of the “x” phase arrives at the “XMux” multiplexer, it switches between identical D_0 values. Note that the “x” phase should still be delayed by the same amount of time as the propagation of signals through the tier one multiplexers. However, unlike the previous multiplexer schemes this delay isn't *critical*. The scheme will still work as long as the rising edge arrives at some time between the edges framing the D_0 bit, and thus the rising edge will not introduce any edges into the next tier. The falling edge of “x” can introduce new edges, but these will not propagate to the final output.

At the final tier, the “ZMux” is used to combine the outputs of the second tier multiplexers. At all times when the select signal is changing, the data at the inputs of the “ZMux” are identical. Therefore, no new edges are introduced. In essence all the dotted potential edges from the first tier appear in the final output, and nothing else.

This new architecture does have the drawback of requiring more phases than the previous schemes, however, the quadrature requirements of the previous architectures and the characteristics of the 4 buffer ring VCO make these available for free. In addition, it should be possible to implement this scheme using just quadrature phases, as the timing for the latter stage multiplexers isn't as critical and can be derived via delays from the existing phases.

Another drawback to this scheme is that generated edges have to pass through 2 levels of multiplexers instead of one. While it is true that this will have negative implications for the signal bandwidth as compared to the simple multiplexer in the original SERDES chip, it is necessary as it alleviates the duty cycle problem to the greatest extent possible. When compared to the SERDES II multiplexer, it has much better bandwidth and subsequent response.

Multiplexer Bandwidth Improvements

The standard CML multiplexer architecture is adequate for most uses, but when the output signal frequency components begin to reach a large fraction of f_T, some of the npn gain has to be traded for bandwidth in order to preserve signal shape. Several types of gain adjustments are being pursued and optimized for.

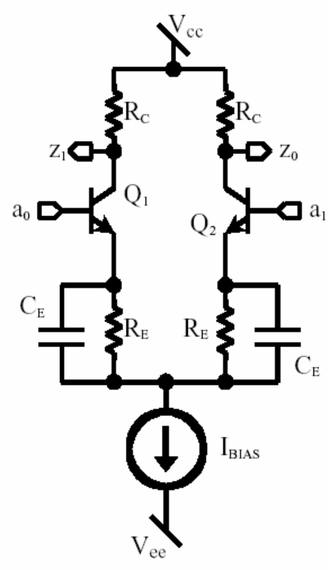

Figure 7.5: Linearized bandwidth improved buffer.

The above buffer uses degenerating resistors to decrease gain for low frequencies. At higher frequencies the emitter capacitors short, causing increased gain at the higher frequencies, effectively increasing the bandwidth. The same technique might be applied to a CML multiplexer as used in the edge-channeling architecture. The presence of the degenerating resistors has a negative impact on the amount of current switched between the two collector resistors RC, which have to be increased to retain the same magnitude output swing as the rest of the gates.

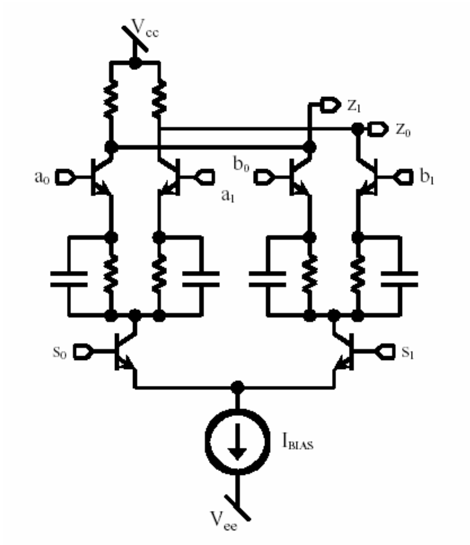

Figure 7.6: Linearized multiplexer.

For a complete linearized multiplexer, however, the resistors and capacitors will have to be placed on the upper level. This will likely not work as the select input switching speed will be greatly affected, and the circuit requires a fast switching rate to ensure proper steering of the edges. There are output capacitor peaking techniques that may also be applicable. It has been found that using the degenerating resistors on the lower select level of the CML multiplexer greatly improves the output characteristics in that there is less distortion caused by the non-linear transitions of the select transistors. This greatly improves the quality of the Edge Steering multiplexer output.

VCO

None of this will be possible without a 10GHz multiphase VCO. Several designs are being pursued, but most of them involve trading off some the wide range of the FFI for more speed. One design in particular uses a “Leap” architecture and a reverse-biased junction for control. The range is small, but the circuit should be fast.

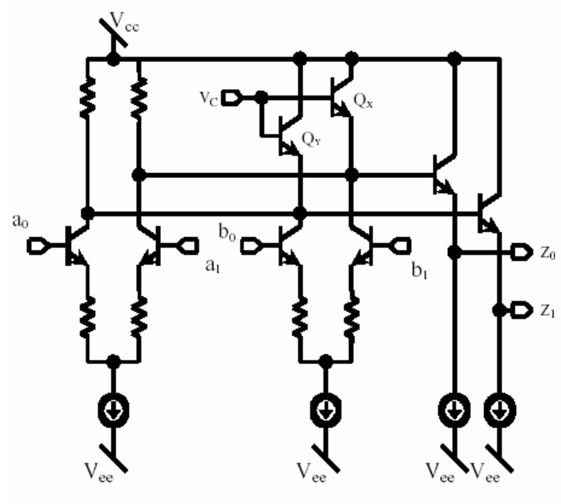

Figure 7.7: New VCO buffer element. Uses fixed leap-forward phase interpolation. Control is achieved via VC reverse-biasing the base-emitter junctions of QX and QY above.

The emitter resistors linearize the buffer so that smooth phase interpolation can occur. The inputs ax and bx, and the zx outputs are all at level two. The control voltage Vc must therefore always remain below the lower level two value of approximately –1.15V and can go as low as –2.9V without exceeding the VBE breakdown. The layout considerations of this VCO are similar to that of previous designs in that a high degree of symmetry is required. In addition, since the connection of phases to the top tier of the edge channeling multiplexer present the only critical symmetrical constraints, the design of the VCO will incorporate the necessary buffers in the proper locations to make the connections to the top tier multiplexers in a symmetrical manner.

Recently, we obtained the opportunity to make use of unused space in a MOSIS 5DM fabrication run. A test VCO as well as supporting circuitry was laid out and submitted within a short deadline period. This VCO has two control inputs, intended for coarse and fine adjustments to the operating frequency. The circuit diagram for one of the buffer elements is included below in Figure 7.8.

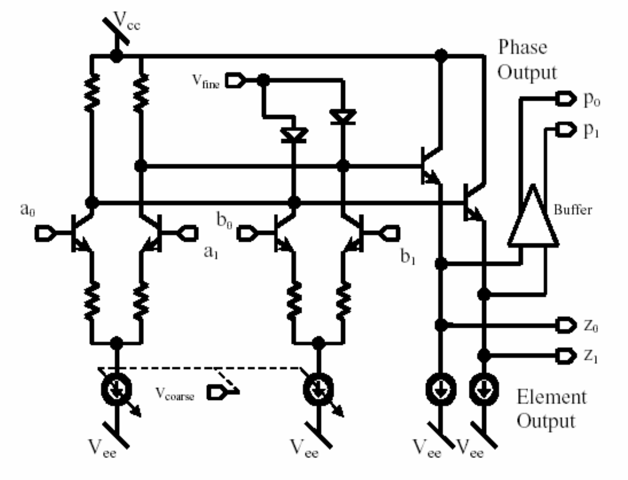

Figure 7.8: Test VCO element fabricated in 5DM.

This delay element features linearized inputs which are evenly interpolated for a 33% speed increase. The coarse control input varies the current in the input trees, affecting the bias. This is done by varying the voltage supplied to the right-half current mirrors. By decreasing the current below that required for maximum f_T, the frequency can be reduced. This is intended to allow us to control the VCO center frequency precisely, and eliminate the center frequency mismatch of previous SERDES designs. Note however, that this modifies the magnitude of the voltage swing developed across the input tree top resistors. This effect is virtually eliminated by the output buffer which also ensures that uneven loading and interconnect doesn't affect the VCO circuit symmetry. The buffer is of the ordinary CML type with emitter followers, giving outputs on level two.

Fine control of the VCO is achieved by the varactors which load the output of the linearized input buffers. They feed a pair of emitter followers which are biased normally, and these in turn drive other element inputs as well as the output buffers.

The VCO was laid out in a highly symmetrical manner as can be seen in Figure 7.9. The majority of the layout was done using only 3 layers of metal. The additional metal layers were used to deliver power evenly and route the fine control line in a symmetrical fashion out through the center of the VCO.

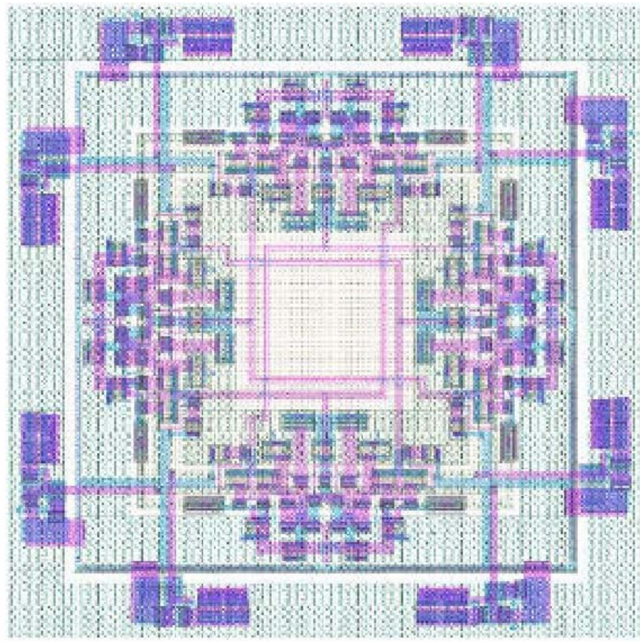

Figure 7.9: 10 GHz test VCO layout.

The voltage references for the current sources at the bases of the current trees are distributed around the perimeter, and tied together, to provide the same voltage for each element. Each element contains an output buffer with connections near the perimeter. It is felt that the layout can be reduced in size by almost a factor of two, but that might prove detrimental to the circuit's performance. Although not easily seen, there are several wired inversions between the delay elements along the centrally routed lines. These inversions, because they are not present on all lines, present an asymmetry. However, if the parasitic effects of the inversions are miniscule with respect to the common parasitics associated with the lines, they should have no real adverse effect on performance. If the layout were to be drastically shrunk, the inversion parasitics would grow proportionally larger and eventually would result in a phase mismatch.

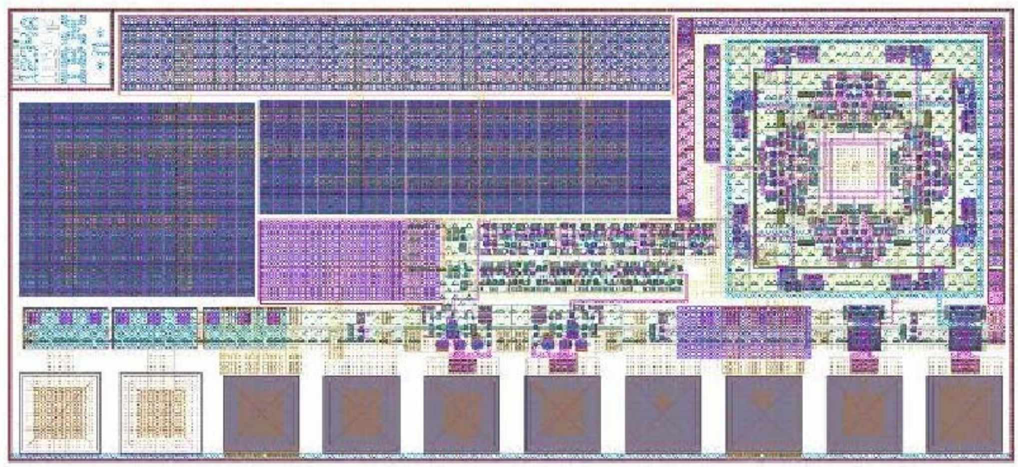

Figure 7.10: Complete 10 GHz VCO test chip layout.

The VCO was embedded in a chip, (Figure 7.10), containing pad drivers and receivers, frequency dividers, and control circuitry for both coarse and fine adjustments. When the chips are delivered we can compare performance to the simulated data below and make adjustments for the next complete SERDES chip, which will be targeted at 40Gb/s.

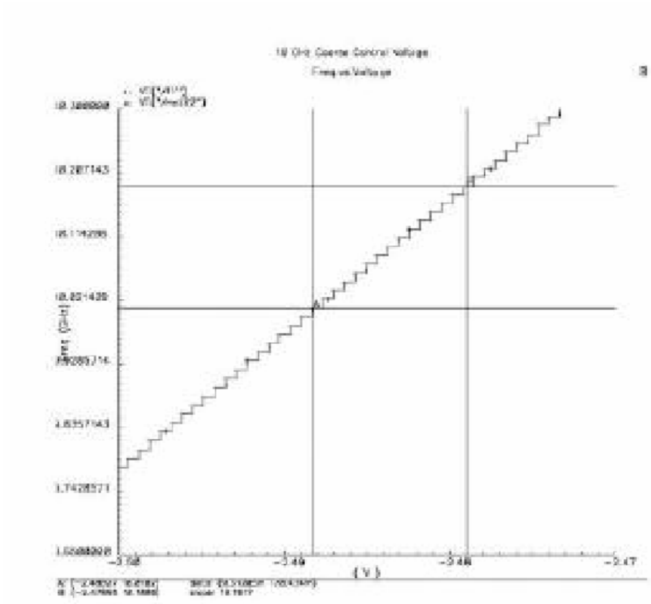

Figure 7.11: Coarse voltage control plot. At center frequency of 10GHz, the control gain is 19.2GHz/V

In Figure 7.11, the VCO frequency is plotted against the coarse control voltage with the fine control voltage fixed at –1.75V. The center frequency is achieved at –2.488V. Note that this coarse control voltage will be regulated by a circuit that will allow a large input swing, causing a small corresponding change in the coarse control voltage. The circuit is basically a modified voltage reference as described previously. An input swing of more than a volt at the chip pad will only cause a few millivolts change in the actual coarse control voltage. The same holds true for the fine control.

Receiver

Clock-data-recovery (CDR) at ratios approaching the goal of 40Gbps in a 50GHz fT have already been demonstrated, therefore the work will focus on the flexibility aspects of modifying loop characteristics for multiple-frequency lock-ins. We plan to investigate a scheme of several local oscillators each running near the desired target frequencies to drive the PLL into lock quickly.

Additional circuitry will be added if we have space to try out various other components that would be useful to have integrated into a SERDES chip. A 16-bit barrel-rotation circuit would be useful from a higher level system perspective to adjust the framing of the bits after the receiver locks. Similarly a parity detector as well as outputs for the detection of various lock and out of lock conditions would be desirable. The CMOS part of the BiCMOS process can be exploited to add various other functional blocks which don't require the raw speed of the HBTs.

New Edge Steering Multiplexer Simulations

The new design has been simulated under a variety of test conditions, ranging from the ideal to the heavily loaded by parasitics. The first simulation results are presented below in Figure 7.12.

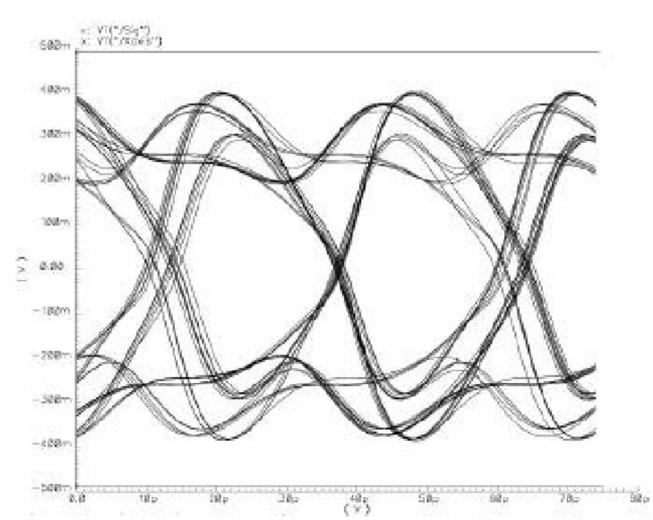

Figure 7.12: Preliminary edge steering multiplexer used in 40Gbps transmitter. The above figure represents a basic design using nominally sized transistors and no optimization.

This shows potential as the circuit was composed of nominal sized transistors embedded in standard CML multiplexers. Delays were not optimized. Encouraged by this result, we undertook the task of improving the characteristics and testing with loads and parasitics. A test bed simulation environment was created to evaluate multiplexer circuits with non-ideal simulated driving circuits with parasitics and loads. Under these conditions, an initial design produced the following output, shown in Figure 7.13.

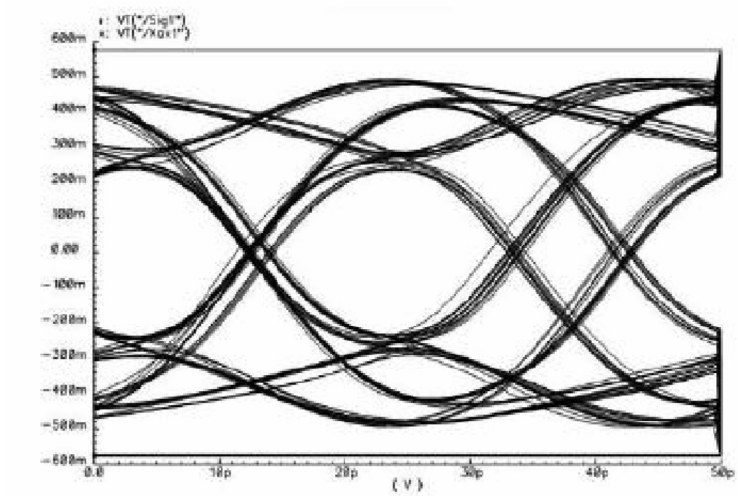

Figure 7.13: Full edge steering multiplexer with non-ideal sources and driving desired load.

As can be seen, the design using plain CML multiplexers shows significant distortion. To investigate causes and solutions, the final two-to-one multiplexer was singled out for first improvements. The final output multiplexer uses 14μm transistors in order to drive a 50Ω load and the external bond pad capacitance. When idealized input was used to drive this multiplexer by itself, the following output resulted.

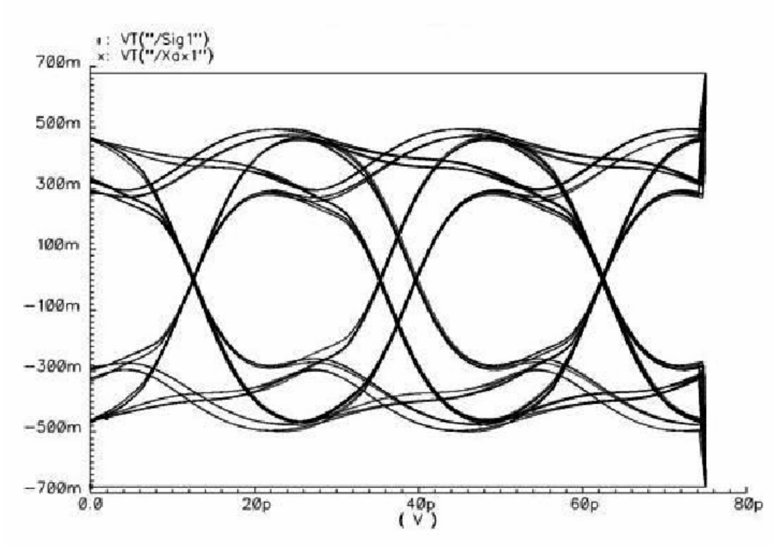

Figure 7.14: Final output multiplexer with idealized inputs driving desired load.

As can be seen, there is a great deal of data dependent jitter present in this output, even though the circuit being tested is only the final multiplexer stage. The effect is occurring because the switch of the lower transistor select pair in a standard CML multiplexer forces a non-linear change in current through the tree. Even though the multiplexer has identical data inputs at the time of the select switch, the output undergoes a significant deviation as the current through the tree is not constant. This deviation from the nominal output value creates data dependent jitter by reducing the transition time of subsequent edges. Several methods have been explored to reduce this effect, including increasing the linearity of the tree current source, and linearizing the CML buffer on the lower level. So far, the latter method appears to gain the greatest improvements in output signal quality. The linearized final output multiplexer output is shown below.

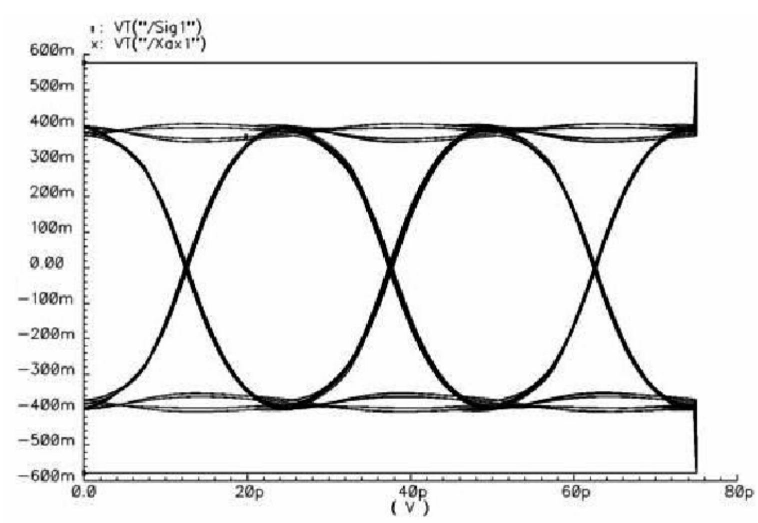

Figure 7.15: Final linearized output multiplexer with idealized inputs driving desired load.

As can be seen in Figure 7.15, the linearization almost eliminates the output ripples. Although this figure uses idealized input from earlier multiplexers, the same techniques are expected to produce a much cleaner output from the multiplexer scheme in general. Currently work is underway to create the optimal edge steering multiplexer design for 5HP. The biases, voltage swings, and transistor sizes must all be individually adjusted and the layout optimized for minimal and matched parasitics. Layout work is expected to be completed in February, and fabrication of the third SERDES design is expected shortly thereafter.

Chapter 8 Conclusion

The DARPA funded SERDES research program at Rensselaer Polytechnic made significant advances in the creation of state-of-the-art high speed circuits in 5HP technology, with benefits which are directly applicable to even high speed circuits in newer technologies such as 7HP. The serializer/deserializer circuits are among the very fastest in terms of data rates for a given technology.

References

[1] D. C. Ahlgren, G. Freeman, S. Subbanna, R. Groves, D. Greenberg, J. Malinowski, D. Nguyen-Ngoc, S. J. Jeng, K. Stein, K. Schonenberg, D. Kiesling, B. Martin, S. Wu, D. L. Harame, and B. Meyerson, “A SiGe HBT BiCMOS Technology for Mixed Signal RF Applications,” BCTM Tech. Dig., pp 195-197 12.3, 1997

[2] R. C. Walker, K. Hsieh, T. A. Knotts, and C. Yen, "A 10 Gb/s Si-Bipolar TX/RX Chipset for Computer Data Transmission," IEEE International Solid-State Circuits Conference, pp. 302-303, 1998.

[3] J. Scheytt, G. Hanke, U. Langmann, “A 0.155-, 0.622-, and 2.488-Gb/s Automatic Bit-Rate Selecting Clock and Data Recovery IC for Bit-Rate Transparent SDH Systems,” IEEE Journal of Solid-State Circuits, vol. 34, no. 12, pp. 1935-1943, December 1999.

[4] D. Friedman, M. Meghelli, B. Parker, H. Ainspan, and M. Soyuer, "Subpicosecond SiGe BiCMOS Transmit and Receive PLLs for 12.5 Gbaud Serial Data Communication," Symposium on VLSI Circuits, pp. 132-135, 2000.

[5] Z. Wang, M. Berroth, J. Seibel, P.Hofmann, A.Hulsmann, K. Kohler, B. Raynor, J. Schneider, “19GHz Monolithic Integrated Clock Recovery Tuning PLL and 0.3um Gate-Length Quantum-Well HEMTs,” 1994 IEEE Int. Solid-State Circuits Conf. (ISSCC), Feb. 1994, Dig. Tech. Pap.,pp. 118-119.

[6] A. Felder, M. Möller, J. Popp, J Böck, H.M. Rein, “46 Gb/s DEMUX, 50Gb/s MUX, and 30 GHz Static Frequency Divider in Silicon Bipolar Technology,” IEEE Journal of Solid-State Circuits, vol. 31, no. 4, pp. 481-486, April 1996.

[7] H. Knapp, T. Meister, M. Wurzer, D. Zoschg, K. Aufinger, L. Treitinger, "A 79GHz Dynamic Frequency Divider in SiGe Bipolar Technology,” IEEE Int. Solid-State Circuits Conf. (ISSCC), Feb. 2000, Dig. Tech. Pap.,pp. 208-209

[8] M. Wurzer, J. Bock, W. Zirwas, H. Knapp, F. Schumann, A. Felder, L. Treitinger, "40 Gb/s Integrated Clock and Data Recovery Circuit in a Silicon Bipolar Technology," IEEE BCTM, 1998, pp. 136-139.

[9] T. Masuda, K. Ohhata, E.Ohue, K.Oda, M.Tanabe, H.Shimamoto, T.onai, K. Washio, "40Gb/s Analog IC Chipset for Optical Receiver using SiGe HBTs," IEEE Int. Solid-State Circuits Conf(ISSCC), Feb. 1998, Dig. Tech. Pap., pp314-315

[10] G. Freeman, M. Meghelli, Y. Kwark, S. Zier, A. Rylyakov, M. Sorna, T. Tanji, O. Schreiber, K.Walter, J. Rieh, B. Jagannathan, A. Joseph, S. Subbanna, "40-Gb/s Circuits Built From a 120-GHz fT SiGe Technology," IEEE Journal of Solid-State Circuits, vol. 37, no. 9, pp.1106-1114, September 2002

[11] G. Georgiou, Y. Beyens, Y. Chen, A. Gnauck, C. Gropper, P. Paschke, R. Pullela, M. Reinhold, C. Dorschky, J. Mattia, T. Mohrenfels, C. Schulien, "Clock and Data Recovery IC for 40-Gb/s Fiber-Optic Receiver," IEEE Journal of Solid-State Circuits, vol. 37, no. 9, pp.1120-1125, September 2002

[12] M. Reinhold, C. Dorschky, E. Rose, R. Pullela, P. Mayer, F. Kunz, Y. Baeyens, T. Link, J. Mattia, "A Fully Integrated 40-Gb/s Clock and Data Recovery IC With 1:4 DEMUX in SiGe Technology," IEEE Journal of Solid-State Circuits, vol. 36, no. 12, pp.1937-1945, December 2001

[13] H. Kroemer, “Heterostructure Bipolar Transistors and Integrated Circuits,” Proceedings of the IEEE, vol. 2, no. 1, pp. 13-25, January 1982
2
[14] C. McAndrew, J. Seitchik, D. Bowers, M. Dunn, M. Foisy, I. Getreu, M. McSwain, S. Moinian, J. Parker, D. Roulston, M. Schroter, P. Wijnen, L. Wagner, "VBIC95, The Vertical Bipolar Inter-Company Model," IEEE Journal of Solid-State Circuits, vol. 31, no. 10, pp.1476-1483, October 1996

[15] Jan M. Rabaey, *Digital Integrated Circuits a Design Perspective*., Englewood Cliffs, NJ:, Prentice Hall, 1996.

[16] R. L. Treadway, “DC Analysis of Current Mode Logic,” IEEE Circuits and Devices Magazine, Volume: 5 Issue: 2, March 1989

[17] David A. Hodges, Horace G. Jackson*, Analysis and Design of Integrated Circuits*., New Yourk,NY: McGraw-Hill, Inc,1988.

[18] Dan H. Wolaver, *Phase-Locked Loop Circuit Design* ., Englewood Cliffs, NJ: Prentice Hall, 1991.

[19] T. H. Lee, and A. Hajimiri, "Oscillator Phase Noise: A Tutorial," IEEE Journal of Solid-State Circuits, vol. 35, no. 3, pp. 326-335, March 2000.

[20] S. A. Steidl, "A 32-Word by 32-Bit Three-Port Bipolar Register File Implemented Using a SiGe HBT BiCMOS Technology," Candidacy document, Rensselaer Polytechnic Institute, Department of Electrical Engineering, May 1999.

[21] S. Finocchiaro, G. Palmisano, R. Salerno, C. Sclafani, “Design of Bipolar RF Ring Oscillators,” IEEE (?) pp. 5-8, 1999

[22] A. W. Buchwald, and K. W. Martin, "High-speed voltage-controlled oscillator with quadrature outputs," Electronics Letters, vol. 27, no. 4, pp. 309-310, February 1991.

[23] H. Matsuoka, and T. Tsukahara, "A 5-GHz Frequency-Doubling Quadrature Modulator with a Ring-Type Local Oscillator," IEEE Journal of Solid-State Circuits, vol. 34, pp. 1345-

1348, September 1999.

[24] P. Kinget, R. melville, D. Long, V.Gopinathan, "An Injection-Locking Scheme for Precision Quadrature Generation," IEEE Journal of Solid-State Circuits, vol. 37, no. 7, pp 845-851, July 2002

[25] S. Lee, B. Kim, and K. Lee, "A Novel High-Speed Ring Oscillator for Multiphase Clock Generation Using Negative Skewed Delay Scheme," IEEE Journal of Solid-State Circuits, vol. 32, no. 2, pp. 1451-1454, February 1997.

[26] S. K. Enam and A. A. Abidi, "A 300-MHz Voltage-Controlled Ring Oscillator," IEEE Journal of Solid-State Circuits, vol. 25, no. 1, pp. 312-315, February 1990.

[27] L. Sun, T. Kwasniewski, and K. Iniewski, "A Quadrature Output Controlled Ring Oscillator Based on Three-Stage sub-feedback Loops," IEEE Internation Symposium on Circuits and Systems, vol. 2, pp 176-179, 1999.

[28] J. A. McNeil, "Jitter in Ring Oscillators," IEEE Journal of Solid-State Circuits, vol. 32, pp. 870-879, June 1997.

[29] G. Niu, Z.Jin, J. Cressler, R. Rapeta, A. Joseph, D. Harame, "Transistor Noise in SiGe RF Technology," IEEE Journal of Solid-State Circuits, vol. 36, no. 9, pp. 1424- 1427, September 2001

[30] T. W. Krawczyk, "Circuits for the Design of a Serial Communication System Utilizing SiGe HBT BiCMOS Technology," PhD Thesis, Rensselaer Polytechnic Institute, Department of Electrical Engineering, Nov 2000.